Studies of *Thermobifida fusca* cellulases Cel9A and Cel48A using a novel kinetic modeling approach for the enzymatic digestion of cellulose.

A Dissertation

Presented to the Faculty of the Graduate School of Cornell University

In partial fulfillment of the Requirements of the Degree of Doctor of Philosophy

Maxim Kostylev

August 2013

Studies of *Thermobifida fusca* Cel9A and Cel48A using a novel kinetic modeling approach for the enzymatic digestion of cellulose.

Maxim Kostylev, Ph.D.

Cornell University 2013

Lignocellulosic biomass is a potential source of sustainable transportation fuels, but efficient enzymatic saccharification of cellulose is a key challenge in its utilization. Enzymatic digestion of cellulose is a complex, heterogeneous process. An enzyme must repeatedly take multiple steps to hydrolyze the substrate to product. In addition, most bulk cellulose contains crystalline, semi-crystalline, and amorphous fractions, whose ratio likely changes during digestion. Access to substrate, rather than hydrolysis, is widely accepted to be the rate limiting step, as we showed experimentally using Cel9A, and the rate of digestion rapidly and continuously drops off as digestion proceeds. Predictive kinetic models typically incorporate different parameters and constants to account for the possible factors responsible for such behavior.

Cellulose hydrolysis by individual cellulases and their mixtures can be modeled with a simple two-parameter model based on a modified classical kinetics scheme. Analogous to a fractal kinetics approach, the specific activity constant is replaced with a time-dependent activity coefficient in order to account for the continuous decrease in the digestion rate. The parameter that quantifies the time dependence of the digestion rate is an intrinsic constant for a given cellulase or mixture on a given substrate. The developed kinetic model was utilized in studies aimed to understand the function of aromatic residues located near the active site tunnel entrance of Cel48A and to develop a distinct model of synergistic cooperation between Cel48A and

Cel9A. Additional experiments were carried out to establish experimentally the catalytic base in family 48 glycoside hydrolases.

BIOGRAPHICAL SKETCH

I was born in the former Soviet Union, in Kishinev (now Chisinau), the capital of Moldavia (now Moldova). In 1991, when I was eleven years old, we immigrated to the United States, right before the collapse of the Soviet Union. My hometown in the US is South Bend, Indiana, where my parents still live.

After finishing high school in South Bend, I attended Indiana University in Bloomington. I initially majored in Biology and added Chemistry as my second major midway through college. After four and half years of college and an additional semester of research in the chemistry laboratory where I had worked as a student, I came to Cornell University to study as a graduate student in the department of Chemistry and Chemical Biology. The main reason I decided to come to Cornell was a well-developed multidisciplinary program on fuel cell research directed by Frank DiSalvo and Hector Abruña. I joined the DiSalvo group, but carried out most work with Bruce van Dover in the Materials Science department. More than three years into the program, with the A examination passed, I decided that even though background in Chemistry and Materials Science would offer me opportunity to do research with many interesting applications, I did not want to pursue a career in these areas. I wanted to go back to Biology, which has always been the science I am most fond of. This was a long decision in the making, but once made, it was a big relief!

I stayed in Cornell and joined the Biochemistry, Molecular and Cell Biology program to work with David Wilson. The second round of graduate work, while strange at first, went in much higher spirits because I was back "in my own plate," as a Russian would say. Five years later, I am writing this sketch, happy to have gone the long way to get here. I am glad I tried Chemistry so that I know it is not something that I want to do. I am glad to have switched so late

in my first Ph.D. program in order to ensure that I can still love what I do now and in the future. I am grateful to have met wonderful advisors in both programs, who have been very supportive throughout my time in Cornell. I hope my experience will make me a better person and coworker down the line.

Next step is a postdoctoral research position in the J. Craig Venter Institute, in San Diego, California.

Родителям

ACKNOWLEDGEMENTS

Many sincere thanks to David Wilson. David is a wonderful person and a great advisor. In my time working with him I felt free to develop my ideas and myself, but whenever I felt lost, a short discussion with David would make me feel that I am back on track. To my friends I often call him an idea generator. David almost always knows what you want to say before you are finished saying it and often has a good advice as soon as you are done talking. He is also extremely patient and very approachable. He never rushed me, but almost always responded within a very short time period to any request. As I go on in my career, I will always think of David as an exemplary scientist.

One of the first people I met in the group was Diana Irwin, who recently passed away. Even though she was retired by that point, she would always make herself available to help me out and would often pinpoint by email the location of almost any random thing in the lab. Thank you, Diana!

Thanks to my dissertation committee members Peng Chen and Linda Nicholson for helpful discussions. Also, special thanks to Linda, with whom I worked as a TA, for setting an example of a great teacher.

Thanks to the current and former labmates – Thu, Jose, Nathan, Tony, Uzma, Alessandra, Eleonora, Ya – for small and big help and the discussions that are very important in any scientific project. I had a chance to work closely with Larry Walker and a few members of his group. Special thanks to Jose Moran-Mirabal, with whom we collaborated on a paper, for teaching me to use the HPLC and getting me started on cellulase assays that became central to my thesis. Also, thanks to Stephane Corgie for many interesting discussions related to work, politics, and life.

I originally came to Cornell to work on a Ph.D. in the department of Chemistry and Chemical Biology, where I worked under the guidance of Frank DiSalvo, Bruce van Dover, and Hector Abruña. Even though I decided that I would like to switch programs, I always had a great relationship with all three advisors and I thank them for being understanding and supportive in my decision to switch.

Friends are important for anything we do in life and this was certainly true here in Cornell. I thank all of the special people who became a part of my life in Cornell. Special thanks to Jo, who became like family to me and helped me go through two knee surgeries.

Thank you, Annie, for making the last couple of months in Cornell so special and fun!

Very special thanks to my parents, who made it possible in many direct and indirect ways for me to love what I do.

Most of my work was funded by the DOE via the BioEnergy Science Center.

Table of Contents

Biographical sketch iii

Acknowledgments vi

Chapter 1: Introduction 1

Kinetic modeling of enzymatic cellulose digestion 3

Synergistic interactions in cellulose hydrolysis 5

Thermobifida fusca Cel9A and Cel48A 19

Thesis organization 21

References 22

Chapter 2: A two-parameter kinetic model based on a time-dependent activity coefficient accurately describes enzymatic cellulose digestion. 34

Abstract 35

Introduction 35

Model Development 36

Methods and Materials 38

Results 42

Discussion 46

References 53

Chapter 3: Determination of the molecular states of the processive endocellulase *Thermobifida fusca* TfCel9A during crystalline cellulose depolymerization 58

Abstract 59

Introduction 59

Methods and Materials 60

Results and Discussion 62

References 70

Chapter 4: Determination of the catalytic base in family 48 glycoside hydrolases 73

Abstract 74

Introduction 74

Methods and Materials 75

Results and Discussion 76

References 78

Chapter 5: Cel48A from *Thermobifida* fusca: structure and site directed mutagenesis of key residues 81

Abstract 82
Introduction 82
Methods and Materials 83
Results and Discussion 87
Conclusions 95
References 103
Chapter 6: A distinct model of synergistic interactions between a processive endocellulase TfCel9A and TfCel48A from *Thermobifida fusca.* 108
Abstract 109
Introduction 109
Methods and Materials 110
Results 114
Discussion 117
References 121
Appendix A (Permission for reuse of published work) 125

CHAPTER 1

Introduction

The need to shift toward a sustainable-energy based economy is becoming generally accepted. The energy used in the US for private and commercial transportation constitutes over a quarter of its total energy consumption (*1*). Liquid fuels are likely to continue being the main form of energy for most forms of transportation while alternative technologies are developed. If produced and utilized efficiently, lignocellulosic biomass is a potential source of renewable, low-carbon-footprint liquid fuels. While the primary form of biofuel in the U.S. is currently corn ethanol, it is widely recognized that a change to cellulose-derived fuel is needed in order to increase energy gain, minimize the carbon footprint, and reduce competition with food. Key challenges associated with such a transition are biomass recalcitrance and enzyme cost. A conventional process for biomass conversion to fuel involves substrate pretreatment to remove lignin and increase accessibility for enzymes, enzymatic digestion of cellulose (and hemicellulose) to fermentable sugars, and microbial conversion of sugars to a combustible fuel. Understanding the mechanisms and properties of cellulases and other enzymes that participate in the digestion of biomass should result in a more effective utilization of these enzymes in the industrial production of cellulosic biofuels and thus reduce the overall energetic and financial costs of the biomass-to-fuel conversion process.

Cellulose digestion by most cellulolytic microorganisms is carried out by cellulases (EC 3.2.1.4) from multiple glycoside hydrolase (GH) families, many of which synergize with each other. In addition, a number of important non-cellulase auxiliary proteins have been recently discovered and studied (*2, 3*). The major function of these proteins appears to be disruption of the recalcitrant portions of cellulose, making them more accessible for digestion by cellulases. As modular proteins, many cellulases contain a catalytic domain (CD) attached to one or more carbohydrate binding modules (CBMs) via a flexible linker. The main function of the CBM is to

increase the effective concentration of the CD on the cellulose surface (*4, 5*). There are two major classes of cellulases. Exocellulases have an active site inside a tunnel and attack cellulose chains from the ends. Most exocellulases appear to be processive and produce cellobiose (G2) as their major digestion product. Endocellulases have an active site inside an open cleft, which allows them to initiate hydrolysis anywhere along the cellulose chain. Most endocellulases seem to have low or no processivity, making random cleavages in accessible regions of the substrate. There is, however, a subclass of processive endocellulases, the most studied of which is *Thermobifida fusca* Cel9A (TfCel9A).

Kinetic modeling of enzymatic cellulose digestion

Digestion of insoluble cellulose involves a number of discrete steps. The enzyme must first bind to the substrate, usually via the CBM. It then must find an accessible site on the substrate and transfer an individual cellulose chain into the active site. Finally, the enzyme hydrolyzes the β-1,4 glycosidic bond, releases the product, and either translates along the chain or releases it. Exocellulases translate along the chain, cleaving off G2 units multiple times before releasing it. Once the cellulose chain is released from the active site, the CD may rebind to the same or a nearby chain or the enzyme may dissociate and rebind elsewhere. Given the fact that cellulase rates of hydrolysis of soluble substrates are much greater than those of insoluble substrates, it is generally believed that access to substrate is the rate limiting step in cellulose digestion. There is now experimental support for this idea (*6*) and we provide additional evidence in Chapter 3, using TfCel9A hydrolysis of BC in the absence and presence of cellotetraose, a competitive substrate for TfCel9A.

Kinetic modeling of enzymatic cellulose digestion is complicated by the number of steps involved and by the heterogeneity of the substrate, which changes during the reaction. It appears

that a constant rate of hydrolysis is never achieved in cellulose digestion and the rate begins to drop off rapidly after the initial digestion (*7*). Various kinetic models of cellulose digestion, some of which are based on the Michaelis-Menten scheme (*8*), have been proposed over the years (as reviewed in (*9*) and (*10*)). Some models include cellulase adsorption, most commonly expressed with the Langmuir isotherm. Incorporation of adsorption is complicated, however, by the fact that only a fraction of the cellulases is productively bound. This productively bound fraction is not directly measurable and most likely continuously changes over time. Nevertheless, despite their apparent limitations, many of the models are able to accurately fit experimental data and predict product formation under various conditions, sometimes up to high extents of digestion. The majority of the proposed models are of a predictive nature, which is particularly useful for the utilization of cellulases on an industrial scale. They typically rely on multiple mechanistic studies that allow for the determination of the relevant parameters included in the models.

Another goal of kinetic modeling is to better understand the mechanisms by which enzymes act on their substrates. The mechanisms and properties of enzymes acting on soluble substrates are commonly deduced using Michaelis-Menten kinetic parameters V_{max}, K_M, and k_{cat}. The Michaelis-Menten kinetic scheme is based on a number of assumptions, which include steady state reaction conditions and a homogeneous substrate. Neither assumption is applicable to the hydrolysis of cellulose, however, which is a heterogeneous material that consists of highly crystalline, semi-crystalline, and amorphous regions (*11*-*13*). While the exact natures of cellulosic substrates have not been thoroughly characterized, due to their complexity and technological limitations, cellulose can be described as a collection of substrates with a spectrum of reactivities, such that some fractions are digested more easily than others. In this scenario one would expect preferential depletion of easily degradable regions early in the digestion followed

by increasingly slower digestion of the recalcitrant fractions of the substrate. A decrease in substrate reactivity with the extent of substrate digestion has been demonstrated (*14, 15*), although the exact reasons for the increased recalcitrance are not easily determined. Some authors also have suggested that enzyme inactivation via irreversible non-productive binding is the primary cause of the continuously decreasing hydrolysis rate of cellulose (*16, 17*).

Synergistic interactions in cellulose hydrolysis

In 1950 Reese et al. (*18*) proposed a mechanism for cellulose hydrolysis, which involved two general components, $C_1$and C_x, acting in sequence. According to the model, the C_1 component first disrupted and swelled the crystalline cellulose possibly releasing soluble oligosaccharides into solution. Then the C_x component, which was shown to have endoglucanase activity, was able to hydrolyze effectively the previously inaccessible substrate along with the soluble oligosaccharides. Furthermore, the activity of the mixture was found to be higher than the activity of each component acting alone, indicating that the components were acting synergistically. In the following years a number of groups began to identify and characterize the specific enzymes present in these components. It became clear that in addition to the endoglucanase activity demonstrated originally for the C_x component, exoglucanase activity was also present in the filtrates of cellulolytic organisms and that the combination of exo- and endo- activities resulted in the synergistic hydrolysis of cellulose (*19-22*). In 1979 Wood and McCrae (*23*) summarized the findings and proposed what became the classical endo-exo model of enzymatic cellulose hydrolysis. According to this model, endocellulases attack the bulk cellulose, creating new chain ends that are susceptible to exocellulase digestion. Exocellulases in turn create more substrate for endocellulases by disrupting the crystalline substrate and/or by exposing previously inaccessible less ordered substrate that is susceptible to

attack by endocellulases. With time it became evident that some exocellulases are also able to act synergistically with each other (*24-27*), giving rise to the suggestion that there are two classes of exocellulases that preferentially attack either the reducing or the nonreducing end of the cellulose chain. This was definitively demonstrated by Barr et al. (*28*) in 1996 using ^{14}O and ^{14}C labeled cellooligosaccharides.

It seems likely that synergism occurs only when two cellulases attack different regions of the cellulose microfibril and that each enzyme creates new sites of attack for other enzymes in the mixture. However, this is an over simplification of a complex process that is still not completely understood. There is no current evidence that synergism requires interactions between the synergizing cellulases, since cellulases from unrelated organisms, which would not have sites for binding to each other, often show cross synergism. However, it is not known if synergizing cellulases interact when they are bound to cellulose and there could be interactions caused by two enzymes binding to a specific site on the cellulose. β-glucosidases, which cleave cellobiose and, to lesser extent, other oligosaccharides (*29*) to glucose, have been shown to enhance cellulase activity (*9, 27, 30*), presumably by relieving product inhibition. Although this is not certain, there is no evidence for other synergistic interactions between β-glucosidases and cellulases. Therefore, these enzymes are not discussed here.

Since the endo-exo model was proposed it has been tested by numerous groups using enzymes from different bacterial and fungal cellulolytic organisms. While most of the data provide support for this model, there is often evidence that the model is not sufficient to describe all aspects of the synergistic interactions between cellulases. It is evident that the methods used to carry out synergistic studies have an impact on the results and their interpretation. Therefore, published data on synergism are often inconsistent and at times contradictory, *despite* the fact

that most experiments support the endo-exo model. This is likely due to the fact that synergism between cellulases involves multiple types of cooperative interactions, which we discuss in this review. In addition, the different factors that affect synergistic experimental data and their interpretation are discussed.

Substrate effect.

The most common pure insoluble substrates used for studying cellulases, listed in the order of increasing crystallinity (which can vary slightly, depending on methods used to measure crystallinity), are phosphoric acid swollen cellulose (PASC), cotton, filter paper (FP), Avicel, and bacterial microcrystalline cellulose (BMCC or BC). All these substrates are heterogeneous and have variable degrees of crystalline order. Due to their complexity it has not been possible to thoroughly characterize the substrates and the most widely used parameter for comparison has been the crystallinity index (CI), which reflects the average value of multiple regions in the substrate. While CI can be a useful parameter for comparing different substrates to each other, it is not a precise or an easily obtainable measurement. CI values depend on the method used to obtain them and variation can be over 30% for the same substrate (*12, 31, 32*). In addition, there is evidence that cellulose may have more than two strictly amorphous and crystalline phases. Using nuclear magnetic resonance spectroscopy Larsson et al. (*13*) demonstrated that various model celluloses contain a significant amount of paracrystalline material, which has intermediate order between pure crystalline and amorphous regions. According to their data, both wood and cotton have over 30% of paracrystalline material. The presence of paracrystalline material along with the specific method chosen to measure the CI may explain why the extent of substrate digestion often does not result in the expected change in CI (*12*). Measured CI values can also partly depend on other substrate properties, such as particle size and surface to volume ratio,

which themselves can affect enzyme behavior. Hence, using changes in CI values during enzymatic digestion to determine enzymatic properties and synergistic mechanisms should always be done with caution (*12, 32*).

Synergistic properties of cellulases are commonly determined from their activities on different substrates when acting alone and together. It has been shown that in a synergistic mixture acting on highly ordered substrate the activities of all components are enhanced (*30, 33*). In (*30*) this was demonstrated by the comparison of processivity of the individual enzymes and their mixtures on FP, while in (*33*) this was done by the pretreatment of bacterial cellulose with one synergistic partner – either an endo- or an exocellulase – before the addition of the other enzyme. However, in another instance, using FP, only pretreatment by endocellulases resulted in the enhancement of the subsequently added exocellulases, while the reverse was not true (*27*). Activities and characteristics of the synergistic mixtures are affected by the substrate used in the studies, as demonstrated by Henrissat et al. (*25*). The authors looked at synergistic combinations of *T. reesei* Cel7A (CBHI), Cel6A (CBHII), Cel7B (EGI), and Cel5A (EGII) on FP, Avicel, homogenized Avicel, BMCC, and *Valonia* microcrystals. For the same total enzyme loads and substrate concentrations, the extent of synergism and optimum enzyme ratios varied substantially between the different substrates. *Valonia* cellulose, which is very highly ordered, showed no synergism at all. Lower synergism on higher ordered cellulose was also observed by Valjamae et al. (*33*) and Jeoh et al. (*34*). In both cases the authors compared the effect of removing amorphous fraction of bacterial cellulose with either acid treatment (*33*) or with digestion by an endocellulase (*33, 34*). They found that the degree of synergism was reduced for the more crystalline material. In some cases pretreatment with acid or with an endocellulase may result in lower synergism between classical endo- and exocellulases due to the decrease of the average

degree of polymerization (DP), as predicted by the functional model of Zhang and Lynd (*35*). Lower DP, however, is unlikely to account for the lack of synergism on *Valonia* cellulose (*25*) or decreased synergism of mixtures that include a processive endocellulase capable of degrading crystalline material (*34*), discussed below.

It is important to note that bulk substrate properties other than crystallinity have a direct impact on hydrolysis rates by individual cellulases and their mixtures. In particular, DP and the total surface area available to cellulases have been shown in various models to affect hydrolysis and, in some cases, the expected synergistic factors (*35-37*). However, these properties specifically affect the kinetic aspects of cellulose digestion, as reviewed in (*9*), and are not likely to affect the mechanisms of synergistic interaction, which are the focus of this review. According to the most recent mechanistic model of enzymatic cellulose hydrolysis to date the placement of the individual cellulose chain into the active site is one of the main limiting steps in the digestion of insoluble substrates (*36*). This is also supported experimentally by high speed atomic force microscopy (AFM) data obtained by Igarashi et al. (*38*). The ability of a given cellulase to access an individual cellulose chain should depend primarily on the mechanistic properties of that cellulase and on the local environment of the chain. Therefore, synergism between cellulases is likely to result, at least in part, from the enhanced access to the individual cellulose chains provided by the cooperative action of different types of cellulases. Given the demonstrated effects of the chosen substrate on synergism, it is likely that better characterization of the heterogeneities present within the commonly used substrates will lead to a more comprehensive understanding of synergistic mechanisms during the digestion of cellulose.

Experimental factors.

In addition to the substrate, many other experimental parameters affect the extent of synergism. These include the specific enzymes in a mixture, the enzyme to substrate ratio, the molar ratios of the enzymes, and the extent of substrate digestion

The endo-exo model of synergism is based on the assumption that there are two distinct classes of cellulases and that some properties are common to all cellulases within either class. Thus, endocellulases are generally presumed to attack at random sites along a cellulose chain, to be nonprocessive, and to attack only the amorphous regions of the substrates. Exocellulases, on the other hand, are presumed to attack only chain ends and are believed to be processive, generating cellobiose as the main product. A number of researchers, however, claim that some exocellulases may have the ability to initiate the attack in an endo- type manner (*23, 25, 39-41*). Regardless of whether some exocellulases are in fact able to initiate hydrolysis by an endo-type attack, it is clear that different exocellulases have different abilities to digest crystalline substrate. For example, while *T. reesei* Cel7A is able to completely digest at least some forms of crystalline cellulose (*42*), *Thermobifida. fusca* exocellulase Cel48A, which is the bacterial equivalent of Cel7A, cannot. In fact, Cel48A is unable to digest more than a few percent of any crystalline substrate (*43*) (and unpublished data). Yet, both Cel48A and Cel7A attack the reducing end of the cellulose chain (*28*), show strong synergism with nonreducing end-directed exocellulases, endocellulases, and processivie endocellulases, and no synergism with each other (*30*). They also make up large fractions of the total secreted cellulases by their respective organisms (*44, 45*). It is not clear whether the mechanism by which Cel48A and Cel7A synergize with other enzymes is the same.

There are also a number of identified processive endocellulases (*46-48*). The most studied of these is *T. fusca* Cel9A, and similar enzymes are present in many cellulolytic bacteria. The catalytic domain (CD) of Cel9A is tightly attached to a family 3c cellulose binding module (CBM3c) (*48*). Based on structural analysis, it is believed that the role of CBM3c is to disrupt the crystalline cellulose surface and then feed the individual cellulose chain to the active site of the CD. This is likely the reason for the relatively high activity of Cel9A on crystalline substrate (*30, 48, 49*). CBM3c is also responsible for the observed processivity of Cel9A, as it provides additional substrate binding sites in line with the substrate binding cleft of the CD (*48-50*). Importantly, Cel9A is able to synergize with all other *T. fusca* cellulases, as well as with *T. reesei* Cel7A and Cel6A (*30*). As a distinct type of a cellulase that combines the properties of classical exo- and endocellulases, Cel9A and other processive endocellulases may have a unique role in synergistic mixtures, which remains to be investigated thoroughly. Other processive endoglucanases, containing GH5 catalytic domains, have been identified in *Saccharophagus degradans,* which mainly produces GH5 cellulases and no known exocellulases (*46*). However, early studies have not shown these enzymes to synergize with other cellulases.

Both, the enzyme to substrate ratio (*51, 52*) and the molar ratios of the enzymes (*25, 34*) in synergistic mixtures are variable parameters that can influence the rate and extent of substrate digestion and thus affect the interpretation of the obtained data. Substrate inhibition of synergistic mixtures has been reported by a number of groups (*52-56*). In one study it was found that *T. reesei* crude cellulases show strong substrate inhibition but that Cel7A, which is the main cellulase in the mixture, was not inhibited (*54*). This suggests that synergistic interactions between enzymes rather than individual enzyme activities were inhibited. It is proposed that at very low enzyme loads the rate limiting step in synergism becomes the dissociation of cellulases

bound unproductively to cellulose and their rebinding to a cleavage site from a synergistic partner. On the other hand, at higher enzyme to substrate ratios, diffusion of enzymes in solution to productive binding sites becomes rate limiting (*52*). The change in apparent rate limiting steps may impact the synergistic behavior of various mixtures of cellulases. Substrate inhibition may be a major cause of the high enzyme load requirement in the industrial production of cellulosic ethanol, as hydrolysis must be carried out at high substrate levels (i.e. 20-40%) to achieve an economical process. This is both, in order to obtain a high volumetric productivity to reduce fixed costs and to obtain high sugar concentrations, which yield high initial ethanol concentrations and reduce subsequent concentration costs.

According to the endo-exo model proposed by Wood and McCrae (*23*) the optimum ratio of endo- to exocellulases should be low, as their function in the model is to create new ends for the exocellulases that carry out most of the hydrolysis. This is consistent with what is seen in nature, as exocellulases make up about 70% of the total cellulase produced by both *T. reesei* (*45*) and *T. fusca* (*44*). In addition, the most abundant enzyme in *Clostridium thermocellum* cellulosomes is an exocellulase (*57*). Experimentally, high optimum exocellulase/endocellulase ratios have been demonstrated in many cases, but not always. Henrissat et al. (*25*) determined optimum ratios of *T. reesei* exocellulase Cel6A and endocellulases Cel5A and Cel7B acting on various crystalline substrates to be about 95:5. On the other hand, the optimum ratio of *T. reesei* Cel7A and the same endocellulases acting on the same substrates was found to be about 1:1. According to the data reported by Jeoh et al. (*34*) the optimum fraction of *T. fusca* endocellulase Cel5A acting together with nonreducing end-directed exocellulase Cel6B on BMCC shifted from 0.1 at low extent of digestion to 0.3-0.5 after about 20% of the substrate was hydrolyzed, as predicted by the functional model of Zhang and Lynd (*35*). However, when Cel5A was acting

together with Cel9A, a processive endocellulase, its optimum ratio remained 0.1 up to 30% hydrolysis. The reasons for such observed differences and what they imply about the mechanisms of synergistic interactions remain unclear. Another surprising result is that for many enzyme pairs there is a broad plateau in the plot of activity versus percent of the added enzyme where the activity is independent of the composition. Given the very different activities of these enzymes this is difficult to explain.

Cellulose binding in synergistic mixtures has been studied by several different groups. Both synergism (*34, 58*) and competition (*34, 59-61*) in binding has been reported. Enhanced binding is usually explained as the creation of new binding sites by the enzymes for each other. Competition is presumed to occur because different enzymes can bind at the same sites on the substrate. This is certainly possible since cellulases usually bind via the CBM, whose binding properties do not correlate with the type of the CD to which it is attached. Another possible explanation, however, is that the apparent decrease in binding is the release of unproductively bound enzymes on the substrate surface by the synergistic interaction of the enzymes. We recently demonstrated that the majority of *T. fusca* Cel9A acting on BMCC is unproductively bound via its CBM2, with an unoccupied CD (*62*). If some fraction of the unproductively bound enzyme is bound irreversibly due to obstacles on the surface of the substrate, a synergistic partner may help release the unproductive enzyme by hydrolyzing that obstacle. This would result in apparent binding competition as measured by the bound fraction of the total added enzyme.

Given the above observations it is likely that the differences and inconsistencies in the reported data on synergism reflect different modes of synergism. In addition to correlating the derived mechanistic explanations of synergism with the specific substrates and enzymes used, it

is important to consider the effect of the extent of substrate digestion on the synergistic interaction. A number of studies have, in fact, illustrated the importance of monitoring synergistic interactions in correlation with the extent of substrate digestion (*33, 34, 60, 63*). Using acid (HCl) hydrolysis or an endocellulase Valjamae et al. (*33*) removed amorphous fractions of bacterial cellulose to various extents before testing the behavior of *T. reesei* Cel7A and Cel7B, alone and together, on the pretreated substrate. Short treatment of cellulose with HCl resulted in reduced activity of Cel7B, but strongly improved the activity of Cel7A. As the treatment time with acid increased, however, Cel7B activity was restored to above the untreated substrate levels, while that of Cel7A was reduced to just above the untreated substrate levels. Both, the net activity and the synergistic factors of the Cel7A/Cel7B mixture decreased with increasing acid treatment time. A similar pattern was observed when the substrate was pretreated with Cel7B instead of acid. All activities were monitored in time course experiments, revealing the different kinetic behavior of the cellulases on pretreated substrates. Based on this data and previously published Monte Carlo kinetic simulations (*64*) the authors make a convincing argument for a novel synergistic mechanism, which is complementary to the classical endo-exo model. According to their model, as the exocellulase processively hydrolyzes individual cellulose chains from the bulk surface, the surface becomes eroded, with obstacles made up of solitary chains that limit the processivity and efficiency of the exocellulases. Endocellulases are able to efficiently remove the solitary chains left on the surface and thus increase the efficiency of the continued substrate hydrolysis by exocellulases. Our time course data using *T. fusca* Cel9A, a processive endocellulase, and Cel48A, a reducing end-directed exocellulase, suggest a similar interaction between these two enzymes (unpublished data). A recent study using high-speed atomic force microscopy to monitor activity of *T. reesei* Cel7A

and Cel6A on highly crystalline cellulose from *Cladaphora* sp. provides additional support for this model (*65*). Continuous monitoring of Cel7A activity alone revealed that the processive enzymes were regularly halted by obstacles on the substrate. In the presence of both Cel7A and Cel6A, however, a much larger fraction of the monitored enzymes was mobile. In another study Josefsson et al. (*63*) looked at real time interaction between several endo- and exocellulases and ultrathin cellulose films using a high resolution quartz crystal microbalance. The physical action of cellulases was monitored from the initial binding to the point where the rate of hydrolysis was almost zero. An important finding in this study was that endocellulases not only create more ends for exocellulases to act on, but that the action of endocellulases appears to significantly swell the film, which would make the new ends much more accessible for the attack by exocellulases. Interestingly, the two endocellulases tested (Cel7B from *T. reesei* and Cel45A from *Humicola insolens*) show different abilities to swell the cellulose film. The possible reasons for these differences, however, were not discussed. Also, only Cel7B was tested in a synergistic mixture with two exocellulases. It would be interesting to determine the effect of individual endocellulase activity in this system on the characteristics of the synergistic mixtures.

Cellulosomes

Many anaerobic bacteria produce large multienzyme complexes called cellulosomes, which play a major role in cellulose degradation (*57, 66*). Most cellulosomal cellulases do not contain a CBM. Instead, the CBM is located on the scaffoldin protein to which the cellulases bind. Cellulosomal cellulases belong to the same families as the free cellulases produced by aerobic bacteria, but GH6 cellulases are not present in most anaerobic bacteria. It has been possible to produce small cellulosomes with a defined structure *in vitro* (designer cellulosomes) and they have been used to demonstrate synergistic activity for cellulosomal cellulases (*67*).

Mini designer cellulosomes were also used to show that most *T. fusca* free cellulases can function in cellulosomes although it is interesting that the GH6 exocellulase was not able to function in the same setup (*67*). The published literature on cellulosomal synergism remains limited, however, and it is unclear whether the physical proximity of the cellulosomal cellulases provides a distinct mode of synergistic interaction. It is possible that the relatively high specific activity of cellulosomes may result from a lack of substrate inhibition (discussed above) as all of the different types of enzymes are close together on a cellulosome.

Synergism on biomass

When grown on cellulose, most cellulolytic bacteria produce hemicellulases, which act to remove hemicellulose bound to cellulose and thus allow the cellulases to access their substrate. In free enzyme systems, many of these hemicellulases contain cellulose binding CBMs presumably to direct the enzyme to cellulose-bound hemicellulose. A number of commonly used biomass pretreatment methods do not remove completely the hemicellulose present in the original biomass (*68, 69*) and it has been shown that the residual hemicellulose can have a negative impact on cellulose hydrolysis by cellulases (*70, 71*). As a result, it appears that hemicellulases are able to act synergistically with cellulases on pretreated biomass, as has been demonstrated in recent studies (*72, 73*). Similarly, cellulosomes from cellulose grown anaerobic bacteria, contain hemicellulases that also act synergistically on biomass substrates (*74, 75*). Additional studies of synergistic interactions between cellulases and other enzymes that act on biomass are likely to have important implications for the industrial production of biofuels, as efficient utilization of such enzymes may significantly reduce costs associated with both biomass pretreatment and enzymatic cellulose hydrolysis.

The "C_1" component.

After the proposal of the C_1 component by Reese et al. (*18*) a lot of effort was invested to identify the function and the enzymes involved in this component. By the time Wood and McCrae proposed the endo-exo model of synergism, this was still an unresolved issue (*23*). Reese and others believed that C_1 disrupts the crystalline surface in order to make the substrate more accessible to endo- and exocellulases (*76*). Wood and McCrae, on the other hand, argued that the C_1 component is composed of exocellulases and that the initial attack on the substrate is carried out by the endocellulases in the C_x component. In their endo-exo model the C_x component carries out the function proposed by Reese for the C_1 component. After it was discovered that cellulases usually have a CBM attached to a CD by an unstructured, flexible linker (*77*), some researchers proposed that the CBM may carry out the role of disrupting the substrate, making it more accessible for the CD (*78-80*). While a number of studies have demonstrated some cellulose disrupting ability by various CBMs from native cellulases, their ability to synergize with CDs appears to be weak at best (*27, 80*).

In recent years, however, a number of disruptive proteins with strong synergistic properties have been discovered. These include a chitin-binding protein CBP21 produced by the chitinolytic bacterium *Serratia marcescens* (*2, 81*), two *T. fusca* proteins E7 and E8 (*82*), CelS2 protein from *Streptomyces coelicolor* (*83*), and family 61 glycoside hydrolases (GH61) produced by plant degrading fungi (*3, 84*). CBP21, E7, E8, and CelS2 are classified in family CBM33 (E8 and CelS2 also contain a family 2 CBM attached to the CBM33). CBP21 has been shown to enhance the activity of chitinases on chitin (*81*), especially in the presence of a reducing agent (*2*). In a similar manner E7, E8, CelS2 and some GH61 proteins enhance the activity of cellulases acting on crystalline cellulose in the presence of various small molecules (*3, 82, 83*).

In two studies of GH61 from *Thermoascus aurantiacus* (*85*) and *Neurospora crassa* (*84*) it was shown that the ability of GH61 to enhance activity of cellulases can also be stimulated by the presence of cellobiose dehydrogenase instead of a small molecule All of the disruptive proteins produce oligosaccharides of various degrees of polymerization and seem to require oxygen for activity. Oxidized products have been detected for CBP21 (*2*), CelS2 (*83*), and GH61 proteins (*85, 86*). The X-ray structures for CBP21 (*87*) and GH61 proteins (*3, 86, 88*) reveal that there are none of the catalytic acid/base residues that are required for cellulase activity. The highly conserved residues in these proteins are found around a metal binding site, which may play a structural and/or a functional role. A number of divalent metal ions have been suggested, but the most thorough study in this regard to date, using GH61 proteins from *T. aurantiacus* provides strong evidence that it is copper (*86*). The authors suggest that GH61 family enzymes should be classified as copper oxidoreductases rather than glycoside hydrolases. While the detailed mechanism of these disruptive proteins remains to be determined, it is generally agreed that they disrupt crystalline cellulose, most likely via oxidoreductive cleavage of the cellulose chains in the bulk crystalline substrate. These proteins thus make cellulose more accessible to attack by cellulases and appear to carry out the function of the C_1 component predicted by Reese et al. (*18*) in 1950.

Perspective

Based on the available information to date, we proposed a model of synergism that incorporates different modes of interactions between cellulose-degrading enzymes (Figure 1.0). This model is based on the idea that cellulose contains a spectrum of crystalline order ranging between strictly amorphous and crystalline material. In this context, it is likely that different enzymes preferentially bind to and possibly act on the different fractions of the substrate. This

may explain why many cellulolytic organisms produce enzymes with apparently redundant activities.

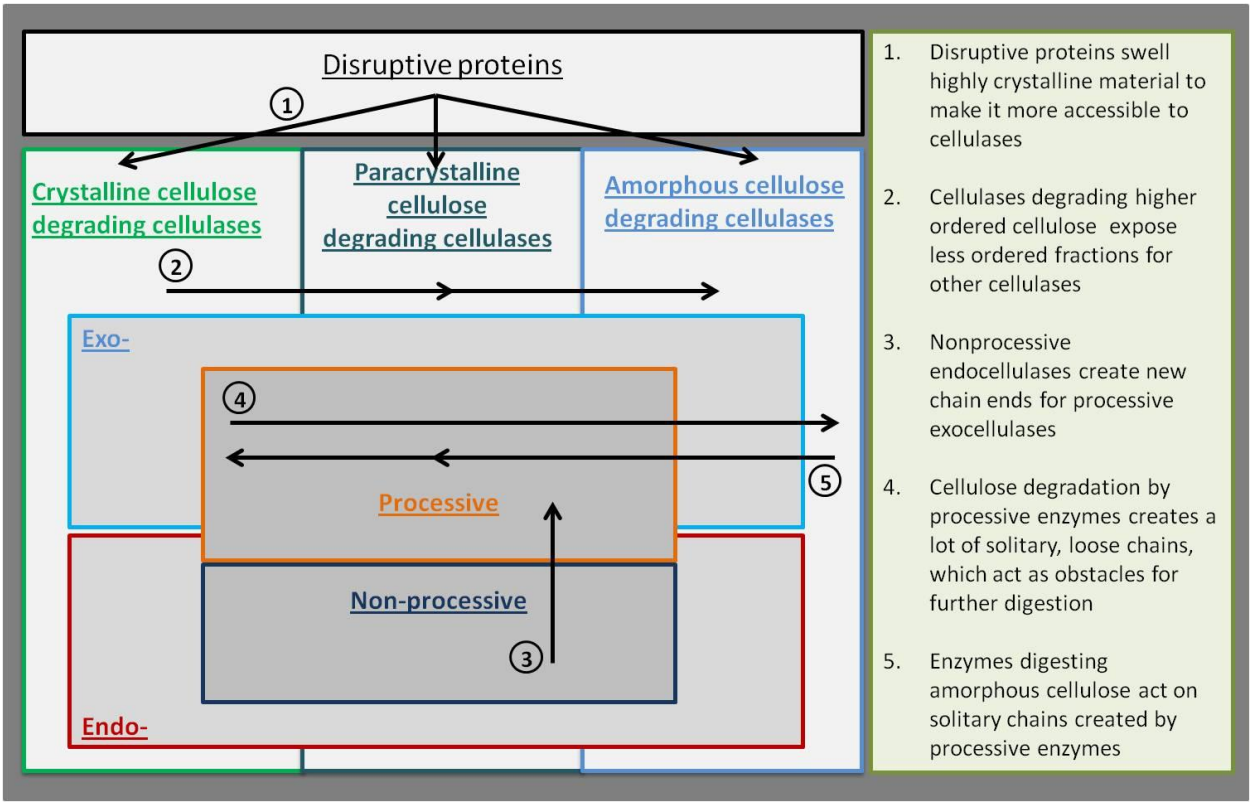


Figure 1.1: Proposed model of synergistic interaction in enzymatic cellulose hydrolysis.

It is generally accepted that the highly ordered fractions of cellulose are responsible for its recalcitrance to complete and efficient digestion, as most cellulases appear to preferentially attack the lesser ordered regions. Given the recent data obtained for proteins from CBM33 and GH61 families, it is likely that these enzymes disrupt the more recalcitrant portions of cellulose and make the substrate more accessible to digestion by cellulases. Efficient digestion by cellulases requires simultaneous action on all parts of the substrate including more and less ordered fractions. An optimum enzymatic mix attacks reducing and nonreducing ends of the chains as well as regions within the chains. As endocellulases attack regions within the chains,

they create new chain ends that are accessible to exocellulases. As processive cellulases – exocellulases and processive endocellulases – digest individual chains, they are likely to create an eroded surface covered by more exposed and thus less ordered chains. The exposed chains may act as obstacles to processive cellulases, but are susceptible to attack by cellulases that are unable to digest more ordered material. Optimum ratios of the different enzymes are likely to depend on the properties of the substrate, especially in the initial stages of digestion.

While the proposed model builds on the classical endo-exo model of synergism, the detailed interactions between its different components are not sufficiently understood and additional mechanistic studies are needed. In Chapter 6 we propose a distinct model of synergism between *T. fusca* Cel9A and Cel48A, which was developed using the kinetic model presented in Chapter 2.

Thermobifida fusca Cel9A and Cel48A

T. fusca is a model cellulolytic bacterium, which has been extensively studied in our group. It secretes six cellulases from four GH families – Cel5A, Cel6A, Cel6B, Cel9A, Cel9B, Cel48A – and two CBM33 (now AA10) disruptive proteins E7 and E8. Most of the work presented in this dissertation is based on assays involving TfCel9A and TfCel48A. TfCel9A is a processive endocellulase, which contains two cellulose binding modules, CBM2 and CBM3c (*50*) and is relatively effective as a single cellulase in the digestion of crystalline cellulose (*30*). CBM2 is attached via a flexible linker (which includes a Fibronectin III domain) to CBM3c. CBM3c is in turn connected almost rigidly to the GH9 catalytic domain (*48*) and is believed to be responsible for the processivity and ability of TfCel9A to digest effectively crystalline cellulose (*48, 50*). TfCel48A is a reducing end directed exocellulase with very poor activity on crystalline cellulose (*43*). Despite its low activity TfCel48A makes up about a third of total

secreted cellulases when *T. fusca* is grown on cellulose (*44*) and it has been shown to interact synergistically with other *T. fusca* cellulases including TfCel9A (*30, 43*). Given that TfCel9A is processive and is the most effective *T. fusca* cellulase on crystalline substrate (*30*), it is unlikely to serve the typical role assigned to endocellulases in the endo-exo model of synergism. In addition, very low activity of TfCel48A on crystalline cellulose also makes it incompatible with the presumed role of exocellulases in the endo-exo model.

Thesis organization

The organization of the thesis is as follows. In Chapter 2 we present a kinetic model of enzymatic cellulose digestion based on a zero order Michaelis-Mented scheme but with a time dependent activity coefficient in place of a constant to account for the decline always observed for cellulose hydrolysis rate. In Chapter 3, using TfCel9A hydrolysis of BC in the presence and absence of G4, we provide experimental evidence that access to substrate is the rate limiting step in enzymatic cellulose digestion. Chapter 4 documents the first experimental data of the catalytic base identity in the GH48 family. One of the TfCel48A catalytic base mutants – D224N – which was prepared for the studies described in Chapter 4, was later crystallized by a collaborating group and the obtained structure is presented in Chapter 5. In the same chapter we also present site directed mutagenesis studies of key residues around the active site tunnel entrance of TfCel48A, which suggest that aromatic residues play an important role in the ability of TfCel48A to access individual chains in crystalline cellulose. Finally, in Chapter 6, we propose a model of synergistic interactions between TfCel9A and TfCel48A, which is distinct from the classical endo-exo model of synergism in enzymatic cellulose hydrolysis. All chapters in the dissertation, including the introduction, are based on published or submitted manuscripts and permission to use published material is provided in Appendix A.

References

1. Administration., U. S. E. I. (2010) Annual Energy Review.
2. Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sorlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides, *Science 330*, 219-222.
3. Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S., and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family, *Biochemistry 49*, 3305-3316.
4. Stahlberg, J., Johansson, G., and Pettersson, G. (1991) A New Model for Enzymatic-Hydrolysis of Cellulose Based on the 2-Domain Structure of Cellobiohydrolase-I, *Bio-Technol 9*, 286-290.
5. Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2012) Determination of the molecular states of the processive endocellulase Thermobifida fusca Cel9A during crystalline cellulose depolymerization, *Biotechnol Bioeng 109*, 295-299.
6. Fox, J. M., Levine, S. E., Clark, D. S., and Blanch, H. W. (2012) Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes, *Biochemistry 51*, 442-452.
7. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., and Westh, P. (2012) Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A, *J Biol Chem 287*, 18451-18458.

8. Michaelis, L., and Menten, M. L. (1913) The kenetics of the inversion effect., *Biochem Z 49*, 333-369.

9. Zhang, Y. H., and Lynd, L. R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems, *Biotechnol Bioeng 88*, 797-824.

10. Bansal, P., Hall, M., Realff, M. J., Lee, J. H., and Bommarius, A. S. (2009) Modeling cellulase kinetics on lignocellulosic substrates, *Biotechnol Adv 27*, 833-848.

11. Kostylev, M., and Wilson, D. (2012) Synergistic interactions in cellulose hydrolysis, *Biofuels 3*, 61-70.

12. Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A., and Johnson, D. K. (2010) Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance, *Biotechnol Biofuels 3*, 10.

13. Larsson, P., Wickholm, K., and Iversen, T. (1997) A CP/MAS13C NMR investigation of molecular ordering in celluloses, *Carbohydr Res 302*, 19-25.

14. Zhang, S., Wolfgang, D. E., and Wilson, D. B. (1999) Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis, *Biotechnol Bioeng 66*, 35-41.

15. Nidetzky, B., and Steiner, W. (1993) A new approach for modeling cellulase-cellulose adsorption and the kinetics of the enzymatic hydrolysis of microcrystalline cellulose, *Biotechnol Bioeng 42*, 469-479.

16. Yang, B., Willies, D. M., and Wyman, C. E. (2006) Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion, *Biotechnol Bioeng 94*, 1122-1128.

17. Eriksson, T., Karlsson, J., and Tjerneld, F. (2002) A Model Explaining Declining Rate in Hydrolysis of Lignocellulose Substrates with Cellobiohydrolase I (Cel7A) and Endoglucanase I (Cel7B) of Trichoderma reesei, *Appl Biochem Biotechnol 101*, 41-60.

18. Reese, E. T., Siu, R. G., and Levinson, H. S. (1950) The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis, *J Bacteriol 59*, 485-497.

19. Eriksson, K. E., and Pettersson, B. (1975) Extracellular enzyme system utilized by the fungus Sporotrichum pulverulentum (Chrysosporium lignorum) for the breakdown of cellulose. 3. Purification and physico-chemical characterization of an exo-1,4-beta-glucanase, *Eur J Biochem 51*, 213-218.

20. Wood, T. M., and McCrae, S. I. (1972) The purification and properties of the C 1 component of Trichoderma koningii cellulase, *Biochem J 128*, 1183-1192.

21. Halliwell, G., and Griffin, M. (1973) The nature and mode of action of the cellulolytic component C1 of Trichoderma koningii on native cellulose, *Biochem J 135*, 587-594.

22. Berghem, L. E., and Pettersson, L. G. (1973) The mechanism of enzymatic cellulose degradation. Purification of a cellulolytic enzyme from Trichoderma viride active on highly ordered cellulose, *Eur J Biochem 37*, 21-30.

23. Wood, T. M., and McCrae, S. I. (1979) Synergism Between Enzymes Involved in the Solubilization of Native Cellulose, 181-209.

24. Fagerstam, L., and Pettersson, L. (1980) The 1.4-β-glucan cellobiohydrolases of Trichoderma reesei QM 9414A new type of cellulolytic synergism, *FEBS Lett 119*, 97-100.

25. Henrissat, B., Driguez, H., Viet, C., and Schülein, M. (1985) Synergism of Cellulases from Trichoderma reesei in the Degradation of Cellulose, *Bio/Technology 3*, 722-726.

26. Tomme, P., Heriban, V., and Claeyssens, M. (1990) Adsorption of two cellobiohydrolases fromTrichoderma reesei to Avicel: Evidence for ?exo-exo? synergism and possible ?loose complex? formation, *Biotechnol Lett 12*, 525-530.

27. Nidetzky, B., Steiner, W., Hayn, M., and Claeyssens, M. (1994) Cellulose hydrolysis by the cellulases from Trichoderma reesei: a new model for synergistic interaction, *Biochem J 298 Pt 3*, 705-710.

28. Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996) Identification of two functionally different classes of exocellulases, *Biochemistry 35*, 586-592.

29. Berghem, L. E., and Pettersson, L. G. (1974) The mechanism of enzymatic cellulose degradation. Isolation and some properties of a beta-glucosidase from Trichoderma viride, *Eur J Biochem 46*, 295-305.

30. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects, *Biotechnol Bioeng 42*, 1002-1013.

31. Park, S., Johnson, D. K., Ishizawa, C. I., Parilla, P. A., and Davis, M. F. (2009) Measuring the crystallinity index of cellulose by solid state (13)C nuclear magnetic resonance, *Cellulose 16*, 641-647.

32. Sathitsuksanoh, N., Zhu, Z. G., Wi, S., and Zhang, Y. H. P. (2011) Cellulose Solvent-Based Biomass Pretreatment Breaks Highly Ordered Hydrogen Bonds in Cellulose Fibers of Switchgrass, *Biotechnol Bioeng 108*, 521-529.

33. Valjamae, P., Sild, V., Nutt, A., Pettersson, G., and Johansson, G. (1999) Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I, *Eur J Biochem 266*, 327-334.

34. Jeoh, T., Wilson, D. B., and Walker, L. P. (2006) Effect of cellulase mole fraction and cellulose recalcitrance on synergism in cellulose hydrolysis and binding, *Biotechnol Prog 22*, 270-277.

35. Zhang, Y. H., and Lynd, L. R. (2006) A functionally based model for hydrolysis of cellulose by fungal cellulase, *Biotechnol Bioeng 94*, 888-898.

36. Levine, S. E., Fox, J. M., Blanch, H. W., and Clark, D. S. (2010) A mechanistic model of the enzymatic hydrolysis of cellulose, *Biotechnol Bioeng 107*, 37-51.

37. Zhou, W., Schuttler, H. B., Hao, Z., and Xu, Y. (2009) Cellulose hydrolysis in evolving substrate morphologies I: A general modeling formalism, *Biotechnol Bioeng 104*, 261-274.

38. Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009) High Speed Atomic Force Microscopy Visualizes Processive Movement of Trichoderma reesei Cellobiohydrolase I on Crystalline Cellulose, *Journal of Biological Chemistry 284*, 36186-36190.

39. Chanzy, H., Henrissat, B., and Vuong, R. (1984) Colloidal Gold Labeling of 1,4-Beta-D-Glucan Cellobiohydrolase Adsorbed on Cellulose Substrates, *FEBS Lett 172*, 193-197.

40. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 A resolution, *Embo J 17*, 5551-5562.

41. Boisset, C., Fraschini, C., Schulein, M., Henrissat, B., and Chanzy, H. (2000) Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the

cellobiohydrolase Cel6A from Humicola insolens and its mode of synergy with cellobiohydrolase Cel7A, *Appl Environ Microbiol 66*, 1444-1452.

42. Chanzy, H., Henrissat, B., Vuong, R., and Schulein, M. (1983) The action of 1,4-beta-D-glucan cellobiohydrolase on *Valonia* cellulose microcrystals. An electron microscopy study., *FEBS Lett 153*, 113-118.

43. Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca, *Eur J Biochem 267*, 4988-4997.

44. Spiridonov, N. A., and Wilson, D. B. (1998) Regulation of biosynthesis of individual cellulases in Thermomonospora fusca, *J Bacteriol 180*, 3529-3532.

45. Ghosh, A., Ghosh, B. K., Trimino-Vazquez, H., Eveleigh, D. E., and Montenecourt, B. S. (1984) Cellulase secretion from a hyper-cellulolytic mutant of Trichoderma reesei Rut-C30, *Archives of Microbiology 140*, 126-133.

46. Watson, B. J., Zhang, H., Longmire, A. G., Moon, Y. H., and Hutcheson, S. W. (2009) Processive endoglucanases mediate degradation of cellulose by Saccharophagus degradans, *J Bacteriol 191*, 5697-5705.

47. Gilad, R., Rabinovich, L., Yaron, S., Bayer, E. A., Lamed, R., Gilbert, H. J., and Shoham, Y. (2003) CelI, a noncellulosomal family 9 enzyme from Clostridium thermocellum, is a processive endoglucanase that degrades crystalline cellulose, *J Bacteriol 185*, 391-398.

48. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca, *Nat Struct Biol 4*, 810-818.

49. Li, Y., Irwin, D. C., and Wilson, D. B. (2007) Processivity, substrate binding, and mechanism of cellulose hydrolysis by Thermobifida fusca Cel9A, *Appl Environ Microbiol 73*, 3165-3172.

50. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of Thermomonospora fusca E4 in cellulose hydrolysis, *J Bacteriol 180*, 1709-1714.

51. Watson, D. L., Wilson, D. B., and Walker, L. P. (2002) Synergism in binary mixtures of Thermobifida fusca cellulases Cel6B, Cel9A, and Cel5A on BMCC and Avicel, *Appl Biochem Biotechnol 101*, 97-111.

52. Valjamae, P., Kipper, K., Pettersson, G., and Johansson, G. (2003) Synergistic cellulose hydrolysis can be described in terms of fractal-like kinetics, *Biotechnol Bioeng 84*, 254-257.

53. Ryu, D. D. Y., and Lee, S. B. (1986) Enzymatic Hydrolysis of Cellulose: Determination of Kinetic Parameters, *Chemical Engineering Communications 45*, 119-134.

54. Ortega, N. (2001) Kinetics of cellulose saccharification by Trichoderma reesei cellulases, *International Biodeterioration & Biodegradation 47*, 7-14.

55. Valjamae, P., Pettersson, G., and Johansson, G. (2001) Mechanism of substrate inhibition in cellulose synergistic degradation, *Eur J Biochem 268*, 4520-4526.

56. Huang, X., and Penner, M. H. (1991) Apparent substrate inhibition of the Trichoderma reesei cellulase system, *Journal of Agricultural and Food Chemistry 39*, 2096-2100.

57. Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides, *Annu Rev Microbiol 58*, 521-554.

58. Jeoh, T., Wilson, D. B., and Walker, L. P. (2002) Cooperative and competitive binding in synergistic mixtures of Thermobifida fusca cellulases Cel5A, Cel6B, and Cel9A, *Biotechnol Prog 18*, 760-769.

59. Kyriacou, A., Neufeld, R. J., and Mackenzie, C. R. (1989) Reversibility and Competition in the Adsorption of Trichoderma-Reesei Cellulase Components, *Biotechnol Bioeng 33*, 631-637.

60. Medve, J., Karlsson, J., Lee, D., and Tjerneld, F. (1998) Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: adsorption, sugar production pattern, and synergism of the enzymes, *Biotechnol Bioeng 59*, 621-634.

61. Ryu, D. D. Y., Kim, C., and Mandels, M. (1984) Competitive Adsorption of Cellulase Components and Its Significance in a Synergistic Mechanism, *Biotechnol Bioeng 26*, 488-496.

62. Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2011) Determination of the molecular states of the processive endocellulase Thermobifida fusca Cel9A during crystalline cellulose depolymerization, *Biotechnol Bioeng*.

63. Josefsson, P., Henriksson, G., and Wagberg, L. (2008) The physical action of cellulases revealed by a quartz crystal microbalance study using ultrathin cellulose films and pure cellulases, *Biomacromolecules 9*, 249-254.

64. Valjamae, P., Sild, V., Pettersson, G., and Johansson, G. (1998) The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model, *Eur J Biochem 253*, 469-475.

65. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttila, M., Ando, T., and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface, *Science 333*, 1279-1282.

66. Bayer, E. A., Lamed, R., White, B. A., and Flint, H. J. (2008) From Cellulosomes to Cellulosomics, *Chem Rec 8*, 364-377.

67. Caspi, J., Irwin, D., Lamed, R., Li, Y., Fierobe, H. P., Wilson, D. B., and Bayer, E. A. (2008) Conversion of Thermobifida fusca free exoglucanases into cellulosomal components: comparative impact on cellulose-degrading activity, *J Biotechnol 135*, 351-357.

68. Alvira, P., Tomas-Pejo, E., Ballesteros, M., and Negro, M. J. (2010) Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review, *Bioresour Technol 101*, 4851-4861.

69. Chandra, R. P., Bura, R., Mabee, W. E., Berlin, A., Pan, X., and Saddler, J. N. (2007) Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics?, *Adv Biochem Eng Biotechnol 108*, 67-93.

70. Bura, R., Chandra, R., and Saddler, J. (2009) Influence of xylan on the enzymatic hydrolysis of steam-pretreated corn stover and hybrid poplar, *Biotechnol Prog 25*, 315-322.

71. Varnai, A., Siika-Aho, M., and Viikari, L. (2010) Restriction of the enzymatic hydrolysis of steam-pretreated spruce by lignin and hemicellulose, *Enzyme Microb Technol 46*, 185-193.

72. Selig, M. J., Knoshaug, E. P., Adney, W. S., Himmel, M. E., and Decker, S. R. (2008) Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities, *Bioresour Technol 99*, 4997-5005.

73. Hu, J., Arantes, V., and Saddler, J. N. (2011) The enhancement of enzymatic hydrolysis of lignocellulosic substrates by the addition of accessory enzymes such as xylanase: is it an additive or synergistic effect?, *Biotechnol Biofuels 4*, 36.

74. Morais, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D. B., and Bayer, E. A. (2010) Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate, *MBio 1*.

75. Morais, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D. B., and Bayer, E. A. (2010) Contribution of a xylan-binding module to the degradation of a complex cellulosic substrate by designer cellulosomes, *Appl Environ Microbiol 76*, 3787-3796.

76. Reese, E. T. (1975), Springer-Verlag, New York.

77. Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and Claeyssens, M. (1988) Studies of the cellulolytic system of Trichoderma reesei QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis, *Eur J Biochem 170*, 575-581.

78. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, A. J., and Kilburn, D. G. (1991) Non-Hydrolytic Disruption of Cellulose Fibers by the Binding Domain of a Bacterial Cellulase, *Bio-Technol 9*, 1096-1099.

79. Teeri, T., Reinikainen, T., Ruohonen, L., Jones, T. A., and Knowles, J. K. C. (1992) Domain Function in Trichoderma-Reesei Cellobiohydrolases, *J Biotechnol 24*, 169-176.

80. Esteghlalian, A. R., Srivastava, V., Gilkes, N. R., Kilburn, D. G., Warren, R. A., and Saddle, J. N. (2001) Do cellulose binding domains increase substrate accessibility?, *Appl Biochem Biotechnol 91-93*, 575-592.

81. Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M., Synstad, B., and Eijsink, V. G. (2005) The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation, *J Biol Chem 280*, 28492-28497.

82. Moser, F., Irwin, D., Chen, S., and Wilson, D. B. (2008) Regulation and characterization of Thermobifida fusca carbohydrate-binding module proteins E7 and E8, *Biotechnol Bioeng 100*, 1066-1077.

83. Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunaes, A. C., Stenstrom, Y., Mackenzie, A., Sorlie, M., Horn, S. J., and Eijsink, V. G. (2011) Cleavage of cellulose by a CBM33 protein, *Protein Sci 20*, 1479-1483.

84. Phillips, C. M., Beeson, W. T., Cate, J. H., and Marletta, M. A. (2011) Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by Neurospora crassa, *ACS Chem Biol*.

85. Langston, J. A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E., and Sweeney, M. D. (2011) Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61, *Appl Environ Microbiol*.

86. Quinlan, R. J., Sweeney, M. D., Lo Leggio, L., Otten, H., Poulsen, J. C., Johansen, K. S., Krogh, K. B., Jorgensen, C. I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C. P., Dupree, P., Xu, F., Davies, G. J., and Walton, P. H. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components, *Proc Natl Acad Sci U S A*.

87. Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H., Eijsink, V. G., and van Aalten, D. M. (2005) Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21, *J Biol Chem 280*, 11313-11319.

88. Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C., and Sandgren, M. (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from Hypocrea jecorina, at 1.6 A resolution, *J Mol Biol 383*, 144-154.

CHAPTER 2

A two-parameter kinetic model based on a time-dependent activity coefficient accurately describes enzymatic cellulose digestion.

Abstract

Kinetic modeling of enzymatic cellulose digestion has been complicated by the heterogeneous nature of the substrate and by the fact that a true steady state cannot be attained. We present a two-parameter kinetic model based on the Michaelis-Menten scheme (*1*), but with a time-dependent activity coefficient analogous to fractal-like kinetics formulated by Kopelman (*2*). We provide a mathematical derivation and experimental support to show that one of the parameters is a total activity coefficient and the other is an intrinsic constant that reflects the ability of the cellulases to overcome substrate recalcitrance. The model is applicable to individual cellulases and their mixtures at low-to-medium enzyme loads. Using biomass degrading enzymes from a cellulolytic bacterium *Thermobifida fusca* we show that the model can be used for mechanistic studies of enzymatic cellulose digestion. We also demonstrate that it applies to the crude supernatant of the widely studied cellulolytic fungus *Trichoderma reesei* and can thus be used to compare cellulases from different organisms. The two parameters may serve a similar role to V_{max}, K_M, and k_{cat} in classical kinetics. A similar approach may be applicable to other enzymes with heterogeneous substrates and where a steady state is not achievable.

Introduction

Kinetic modeling of enzymatic cellulose digestion is complicated by the number of steps involved and by the heterogeneity of the substrate, which changes during the reaction. Developed here is a simple two-parameter fit, which is based on the Michaelis-Menten scheme, but with a time dependent activity for the enzyme. This is analogous to the fractal and fractal-like kinetics approach formulated by Kopelman (*2*) for heterogeneous reactions limited by surface diffusion. The objective for this model was to develop a mathematical relationship, which yields an intrinsic constant that quantifies the different abilities of cellulases to overcome

substrate recalcitrance. The two parameters are determined from time course data of cellulose hydrolysis and can provide insight into the mechanisms by which enzymes digest insoluble cellulose, alone and in mixtures. This is best illustrated by comparing the parameters of mutated versions of a cellulase, TfCel48A, with its native form. In addition, we demonstrate how addition of an auxiliary protein to an exocellulase affects its measured parameters, and thus confirms the presumed role of auxiliary disruptive proteins in cellulose hydrolysis. The model applies to low-to-medium enzyme-substrate ratios, and the two parameters can serve a similar role as V_{max}, K_M, and k_{cat} in classical kinetics. In this study we mostly used enzymes from a model cellulolytic bacterium *T. fusca* to validate the proposed model and to demonstrate its utility, but we also provide data for the crude supernatant of a commonly studied cellulolytic fungus, *Trichoderma reesei*. We believe that the demonstrated modeling approach may prove useful for other enzymes that utilize a heterogeneous, continuously evolving substrate and do not reach steady state.

Model Development

In classical kinetics, the reaction rate is constant throughout the reaction. This is also true for the specific activity of enzymes acting on a soluble, homogeneous substrate, before other factors such as substrate depletion, thermal inactivation, and product inhibition become significant. In fractal and fractal-like systems, however, if the reaction is limited by surface diffusion of the reactants, the reaction rate is not constant, but is dependent on time, as demonstrated by Kopelman (*2*):

$$k = Kt^{-h} \qquad 0 \leq h \leq 1 \qquad (2.1)$$

where *t* is time, K is the initial rate coefficient at $t = 0$, and *h* is the fractal factor, which can be determined experimentally and, in some cases, predicted mathematically.

Cellulose digestion can be represented as:

$S \rightarrow P$

where S is insoluble cellulose and P any soluble oligosaccharide. In a classical kinetic system the rate of product formation at low enzyme (*E*) loads (*E* << Substrate) can be represented as:

$$\frac{dP}{dt} = kE_{tot} \quad (2.2a)$$

where *k* is the specific activity of the enzyme. Total generated product (P_{tot}) at any given time can be obtained by integrating Eqn. 2.2a with respect to time:

$$P_{tot} = kE_{tot}t \quad (2.2b)$$

It was determined experimentally that cellulose digestion shows the same time dependence of the specific activity coefficient as described for fractal systems. In addition, only some fraction of the added enzyme (E_{tot}) is productively bound (E_P), such that:

$$E_{tot} = E_P + E_U$$

where E_U is unable to generate a soluble product either because it is not bound or is bound unproductively. Most likely E_P changes continuously during the reaction, but it is not directly measurable. Accounting for the time dependence of the specific activity of the cellulase and considering only the productively bound fraction gives:

$$P_{tot} = KE_P t^{(1-h)} \quad (2.3)$$

It is also not possible to determine the exact value of *K* (i.e. the inherent specific activity) of any cellulase for its activity on insoluble substrate. Thus, for fitting purposes, the two variables, K and E_P, are lumped into one term *A*, which represents the net activity of the total enzyme acting on the substrate, as determined by soluble products formed over time. While it has no effect on the utility of the model, it is important to note that for non-processive cellulases the value of *A* may be an underestimate of the actual number of cleavages carried out by these enzymes because in theory it should take multiple random cleavages to produce a soluble oligosaccharide product.

However, pseudoprocessivity – a repeated non-processive attack in the same region, which produces soluble oligosaccharides – appears to be common among non-processive endocellulases, as evidenced by constant ratios of soluble and insoluble reducing ends generated by such enzymes over time. For the sake of simpler representation and to distinguish the present equation from fractal-like kinetics (discussed below), (1-*h*) is also simplified to a parameter *b*, which we call the "hydrolysis power factor":

$$P_{tot} = At^b \qquad 0 \leq b \leq 1 \tag{2.4}$$

When $b = 1$, the specific activity of the enzyme is constant over time (extent of digestion) as would be the case in classical kinetics. It is useful to monitor cellulase activity with respect to substrate digestion. Knowing P_{tot} allows one to calculate directly the fraction of the digested substrate (X):

$$X = \frac{P_{tot}}{S_{tot}} \times 100\% \tag{2.5}$$

where S_{tot} is the total substrate added to the reaction. Most of the data and empirically determined parameters in the manuscript are presented with respect to X, such that

$$X = At^b \tag{2.6}$$

where A contains S_{tot}.

Materials and Methods

Substrates. Bacterial cellulose (BC) was a gift from Monsanto. BC cake was washed three times with deionized (DI) water by centrifugation and resuspended in DI water with 0.04% sodium azide (Sigma-Aldrich). Concentration was determined as dry weight per volume. Avicel® powder (PH-105; FMC Corporation, Philadelphia, PA, USA) was suspended in DI water with 0.04% sodium azide. Phosphoric acid swollen cellulose (PASC) was prepared from

Avicel® powder using procedures described in (*3*) and was stored in DI water with 0.04% sodium azide.

Enzymes. All individual cloned *T. fusca* cellulases (except TfCel48A) and E7 were expressed in *Escherichia coli* BL21 or *Streptomyces lividans* strains, which secrete these enzymes during expression. The enzymes were purified from the culture supernatant as previously described (*4*, *5*).

TfCel48A contains a 6-His tag on both the amino- and carboxyl termini and no signal peptide. WT and mutant TfCel48A were expressed inside *E. coli BL21*-CodonPlus® (*DE3*)-*RIPL* cells (Agilent Technologies) as follows. A starter culture was grown in Lysogeny Broth (LB) medium overnight at 37 °C. The starter culture was diluted 33 times in fresh LB and the cells were grown at 37 °C until optical density at 600 nm reached ~0.8 (about 2.5 hrs). The culture was transferred to 25 °C and Isopropyl β-D-1-thiogalactopyranoside was added to 0.8 mM. The culture was incubated with shaking for 16 hours. Cells were harvested by centrifugation and resuspended in 1/20 original culture volume of 20 mM sodium phosphate buffer, pH 8.0, 0.01 M Imidazole (Sigma-Aldrich), 0.5 M sodium chloride (Solution A). Cells were lysed with a French Press (at cell pressure of 20,000 PSI). Lysed cells were placed in a 50 °C water bath for 30 minutes (to precipitate out *E. coli* proteins) and centrifuged to remove the precipitate. The supernatant was diluted in 5 additional volumes of Solution A and loaded on a nickel Sepharose 6 Fast Flow (GE Healthcare) column by gravity. The protein was eluted by a gradient of 0.01 – 0.5 M Imidazole in 20 mM sodium phosphate buffer, pH 8.0, 0.5 M sodium chloride. Fractions containing TfCel48A, as determined by gel electrophoresis, were combined and diluted 10-15 times with DI water (to lower the conductivity of the solution to less than 0.5 mA). The protein was then loaded on a Q Sepharose (Sigma-Aldrich) column equilibrated with

0.1 M sodium chloride in 5 mM Bis-Tris buffer, pH 5.8, 10% glycerol (v/v). The protein was eluted by a gradient of 0.1-0.5 M sodium chloride in 5 mM Bis-Tris buffer, pH 5.8, 10% glycerol (v/v). Fractions containing TfCel48A, as determined by denaturing gel electrophoresis, were combined and concentrated using Millipore® centrifugal filter units with a 30 kDa cutoff membrane. Buffer exchange was carried out in the same filter units by washing the concentrated protein three times with 5 mM sodium acetate buffer, pH 5.5, 10% glycerol (v/v). Protein was filtered through a 0.22 μm filter using a syringe and stored at -20 °C.

All enzyme concentrations were determined by spectroscopy using NanoDrop® 1000 spectrophotometer (See Table 2.1 for relevant extinction coefficients and molecular weights). *T. fusca* crude supernatant was obtained by growing *T. fusca* ER1 strain on cellulose for 72 hours, as described in (*6*). *T. reesei* crude supernatant was obtained by growing *T. reesei* L27 strain on cellulose for 96 hrs, as described in (*7*). Cells were removed by centrifugation and the supernatant was stored at -70 °C. Crude supernatant protein concentration was determined by Bradford assay using Bio-Rad Protein Assay Dye Reagent Concentrate and bovine serum albumin for reference. TfCel48A mutants were generated using Agilent Quickchange® II XL Site Directed Mutagenesis Kit following the manufacturer's instructions. All mutations and sequences were verified by Sanger sequencing. Mutant proteins were expressed and purified in the same way as WT.

Time course assays. All reactions were conducted in triplicate in Eppendorf 2 mL Protein LoBind plastic tubes. 1 mg substrate was combined with an indicated amount of enzymes in 0.6 mL 50 mM sodium acetate buffer, pH 5.5. In reactions involving E7, 1 mM glutathione (Sigma-Aldrich) was added to enhance activity of E7 (*8*). Only buffer and substrate were combined for negative controls. Upon mixing, BC and PASC reactions were immediately placed in a 50 °C

Table 2.1: Molecular weights and extinction coefficients of the *T. fusca* cellulases and the disruptive protein E7.

Protein	Molar extinction coefficient (L mol^{-1} cm^{-1})[1]	Molecular weight (kDa)[1]
Cel5A	97,100	46.3
Cel6A	81,800	43.0
Cel6B	115,150	59.6
Cel9A	210,670	90.4
Cel9A-CD	165,480	68.0
Cel9B	207,840	101.2
Cel48A	260,300	105.7
Cel48A-CD	212,610	71.8
E7	70,180	21.3

[1] Values were obtained with DNASTAR® Protean 8.0 software.

(unless otherwise indicated) water bath. Avicel® reactions were carried out in a 50 °C incubator with rotation to prevent the substrate from settling. Samples (in triplicate) were removed at the given time points and placed on dry ice to stop the reaction. Frozen samples were later placed in a boiling bath for 10 minutes in order to denature the enzyme. It was verified experimentally that boiling does not alter soluble sugar profiles detected by High Performance Liquid Chromatography (HPLC). The remaining substrate was removed using Corning® Spin-X® Centrifuge tube filters and the soluble sugar concentrations were measured using a Shimadzu HPLC system fitted with a Bio-Rad Aminex® HPX-87P analytical column and a refractive index detector. The mobile phase was Milli-Q water at a flow rate of 0.6 ml/min. Sample injection (50 µl volume) was performed by an autosampler installed on the instrument.

Apparent K_M assays. 75 nM TfCel48A was combined with the indicated amount of BC in 0.6 ml of 50 mM sodium acetate, pH 5.5 and reactions were incubated in a 50 °C water bath for one hour. Samples were boiled for five minutes to denature the enzyme and were processed and analyzed by HPLC in the same manner as described for the time course assays.

Data analysis. HPLC data were processed with OriginPro 8. Product identities and concentrations were determined by Gaussian peak fitting, using standard solutions with known concentrations of soluble cellooligosaccharides for reference. Soluble sugar concentrations at

time zero were subtracted from all of the subsequently obtained concentrations. Soluble sugar produced upon initial mixing of the enzyme and the substrate is primarily due to the burst activity, as described in (*9*, *10*), whereas the model presented here is concerned with the digestion of the more recalcitrant portions of cellulose. Apparent K_M and the *A* and *b* parameter values of time course profiles were determined using the nonlinear least squares fit of the Michaelis-Menten expression ($V_{max}[S]/(K_M + [S])$) and Eqn. 2.6 respectively. OriginPro 8 default settings (chi-square minimization) were used for all fits.

Results

When the activity of different cellulases is monitored over time, their time course profiles are distinct (Fig. 2.1a). This is also true for the same cellulase acting on different substrates (Fig. 2.1b). It was empirically determined that initial time course profiles of all *T. fusca* cellulases follow Eqn. 2.6, a two parameter expression based on a pseudo zero order (with respect to the substrate) reaction with a time-dependent activity coefficient (Table 2.2). The same equation applies to mixtures of enzymes, as shown for *T. fusca* and *T. reesei* crude supernatants acting on bacterial cellulose (BC) and Avicel® (Fig. 2.1c and d).

Based on the derivation of Eqn. 2.4, the two parameters quantify the net activity of the added cellulase (*A*) and the intrinsic ability of the cellulase to overcome substrate recalcitrance (*b*, the curvature of the time course profile). In order to verify the mathematically implied significance of the parameters experimentally the effect of reaction temperature (Fig. 2.2) and enzyme concentration (Fig. 2.3) on their values was tested. *A* and *b* values were determined for a processive endocellulase TfCel9A on BC at temperatures between 10 and 50 °C. An increase in the reaction temperature had a strong effect on the measured value of *A.* The relationship between *A* and temperature follows the Arrhenius equation (Fig. 2.2b), which holds true for most

reaction rate constants. On the other hand, the value of *b* increased only slightly with temperature, indicating that temperature plays a minor role in the enzyme's ability to overcome substrate recalcitrance. The effect of enzyme concentration on the *A* and *b* values was determined for a processive endocellulase TfCel9A, a non-processive endocellulase TfCel5A, and *T. fusca* crude supernatant on BC (Fig. 2.3). For both individual cellulases, *A* values increased with increasing enzyme concentration. On the other hand, the *b* value was constant, consistent with it being an intrinsic constant. For crude supernatant, the results are similar to those obtained for individual cellulases, but there is a slight increase in the value of the *b* parameter from 0.63 ± 0.01 to 0.69 ± 0.01 with increased enzyme load. This may be due to an improved synergistic effect caused by more densely crowded cellulases on the cellulose surface (*11*).

In recent years, a number of disruptive enzymes, which are important for enzymatic biomass digestion, have been discovered (*8, 12*). These enzymes belong to auxiliary activity families AA10 (formerly CBM33) in bacteria and AA9 (formerly GH61) in fungi, and use an oxidoreductive mechanism rather than general acid/base hydrolysis to cleave cellulose chains at random locations on the bulk substrate (*8, 13, 14*). Their activity is believed to disrupt the crystalline regions of cellulose and thus make it more accessible to cellulases. Addition of these enzymes to cellulases acting on cellulose can significantly increase the rate of cellulose digestion even though they produce very few soluble products by themselves. *T. fusca* secretes two known AA10 enzymes – E7 and E8 – which have been shown previously to stimulate the activity of other *T. fusca* cellulases on BC (*15*). In order to determine the effect of E7 addition on the *b* parameter of an individual cellulase, time course experiments were carried out on BC using *T. fusca* exocellulase TfCel48A in combination with E7 (Fig. 2.4). TfCel48A has the lowest activity

of all *T. fusca* cellulases on insoluble cellulose (Table 2.2). Due to E7's extremely low generation of soluble products and different mechanism of hydrolysis, it is not possible to fit its time course data with the proposed model. When E7 is added to TfCel48A, the *b* parameter of the digestion curve is increased from 0.34 ± 0.01 for TfCel48A alone to 0.65 ± 0.01 for the mixture of the two enzymes. The *A* parameter is increased from 0.53 ± 0.02 to 0.70 ± 0.01.

Table 2.2: *A* and *b* values of all *T. fusca* cellulases digestion time course profiles on BC.

Enzyme[*]	Endo/ Exo	$A^{\#}$ (% digestion hr^{-1})	$b^{\#}$	Processivity[%]
TfCel9A	Endo	1.68 ± 0.04	0.58 ± 0.02	6.9
TfCel9A-CD	Endo	0.90 ± 0.01	0.58 ± 0.01	2.5-3.5
TfCel5A	Endo	1.61 ± 0.02	0.33 ± 0.02	N/A
TfCel9B	Endo	1.11 ± 0.03	0.31 ± 0.01	N/A
TfCel6A	Endo	1.18 ± 0.07	0.38 ± 0.01	N/A
TfCel6B	Exo	0.48 ± 0.01	0.50 ± 0.01	20
TfCel48A	Exo	0.52 ± 0.01	0.33 ± 0.01	15
TfCel48A-CD	Exo	0.24 ± 0.01	0.33 ± 0.01	15
T. fusca crude supernatant	N/A	13.97 ± 0.23	0.69 ± 0.01	N/A
T. reesei crude supernatant	N/A	15.32 ± 0.22	0.64 ± 0.01	N/A

[*]Individual cellulase concentrations, except TfCel48A, were 33.3 nM; TfCel48A and TfCel48A-CD concentrations were 100 nM; *T. fusca* and *T. reesei* crude supernatant total protein added per reaction was 95 µg and 8 µg respectively.
[#]N = 3 for all data points. Averaged values with standard deviations were fit to Equation 6 to obtain *A* and *b* values. Standard error is provided for all values. All $R^2 > 0.97$.
[%]Values for TfCel9A are the ratio of soluble and insoluble reducing ends after digestion of filter paper and were previously published (*5, 16*). Values for TfCel6B and TfCel48A are the ratio of cellobiose to cellotriose concentrations (*17*) after digestion of BC. N/A = not applicable.

Access to substrate is widely accepted to be the rate limiting step of enzymatic crystalline cellulose digestion (*18, 19*) and it is likely that cellulases have evolved residues that play a direct

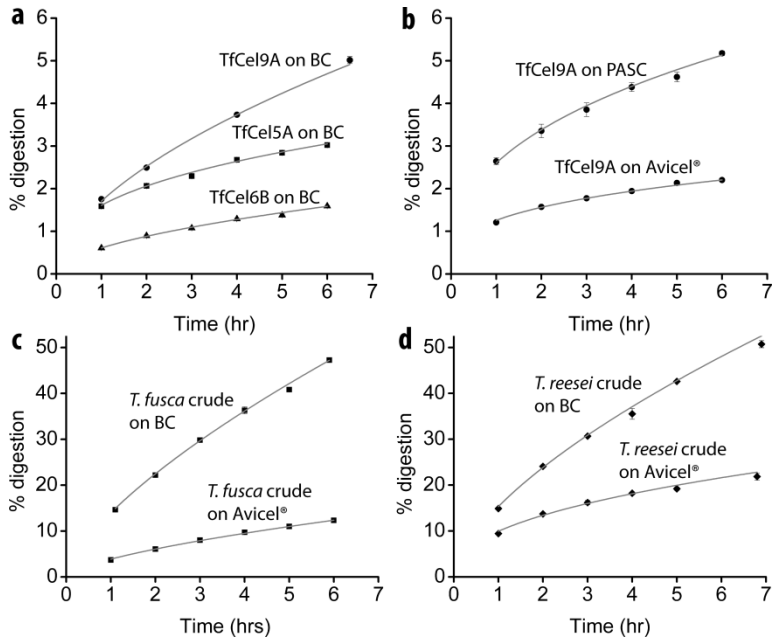


Figure 2.1: Fitted digestion curves for different enzymes on various substrates. (a) 33.3 nM T. fusca Cel9A, Cel5A, and Cel6B on BC; (b) T. fusca Cel9A on Avicel® (167 nM enzyme) and phosphoric acid swollen cellulose (PASC; 33.3 nM enzyme); (c) T. fusca crude supernatant on BC (95µg tot. protein/mg substrate) and Avicel® (159 µg tot. protein/mg substrate); (d) *T. reesei* crude supernatant on BC and Avicel®.

role in the separation of an individual cellulose chain from the bulk substrate. The mutation of such residues should decrease the activity of a cellulase on crystalline cellulose. As for all exocellulases, the active site of TfCel48A is located inside a tunnel (*20*), which is presumed to be important for the processivity of these enzymes. Our group is in the process of studying residues around the tunnel entrance of TfCel48A in regard to their role in crystalline cellulose digestion. As part of this study, the effect of these residues on the *b* parameter value of the enzyme was determined. Fig. 2.5b shows time course data for two TfCel48A mutants, W313A, and W315A on BC. Both residues are conserved in the GH48 family and aromatic residues of various cellulases have been shown to play an important role in substrate binding and cellulase processivity (*17, 21-23*). W313 and W315 are located at the mouth and inside the tunnel respectively (Fig. 2.5a). Both W313A and W315A mutants have strongly reduced activity on BC, as shown by their time course profiles. Both mutants have lower than WT processivity (Fig.

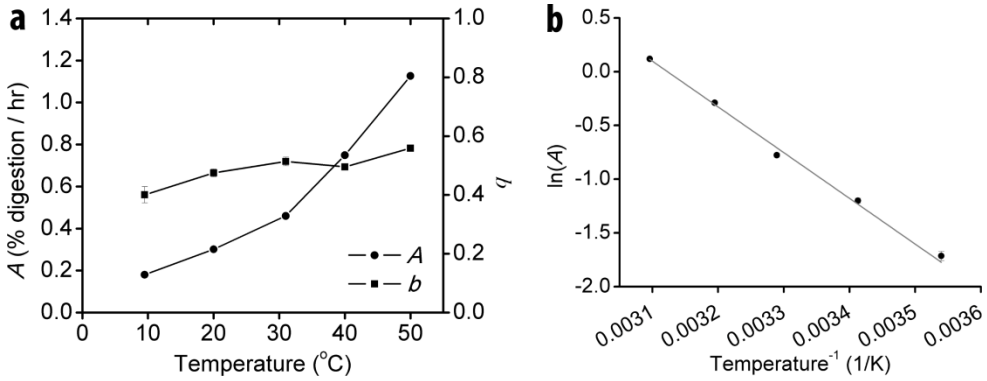


Figure 2.2: Dependence of *A* and *b* on temperature. (a) Experimentally determined *A* and *b* values for TfCel9A digestion of BC at different temperatures; (b) Arrhenius plot of *A*. A time course profile was obtained for each temperature (N=3) and the data were fit with Eqn. 2.6 to obtain *A* and *b* values. Error bars represent standard error of the fit.

2.5d), which is determined by their cellobiose to cellotriose ratio (*17*), and a similarly reduced *A* parameter value. However, only the *b* parameter of W313A is significantly reduced in comparison to WT. A similar result is observed for apparent K_M of the mutants on BC (Fig. 2.5c). The mutation of W315 results in a slight increase in apparent K_M of the enzyme from 1.8 ± 0.3 to 2.4 ± 0.1 mg/ml BC, but the mutation of W313 results in a much greater increase of apparent K_M to 4.4 ± 0.4 mg/ml BC. These results are consistent with the relative position of the two residues: one at the mouth and one inside the tunnel, as only residues at or near the surface of the protein should directly affect the enzyme's interaction with the bulk substrate.

Discussion

The distinct time course profiles of different cellulases acting on the same substrates or the same cellulase acting on different substrates (Fig. 2.1) indicate that the detailed change in substrate during digestion is not universal, but is specific to each cellulase. Hence, the time course profile is a reflection of the ability of a cellulase to digest various forms of cellulose. Enzymatic cellulose digestion is a multistep process, which is further complicated by the continuously changing substrate. Most models of cellulose hydrolysis thus utilize multiple

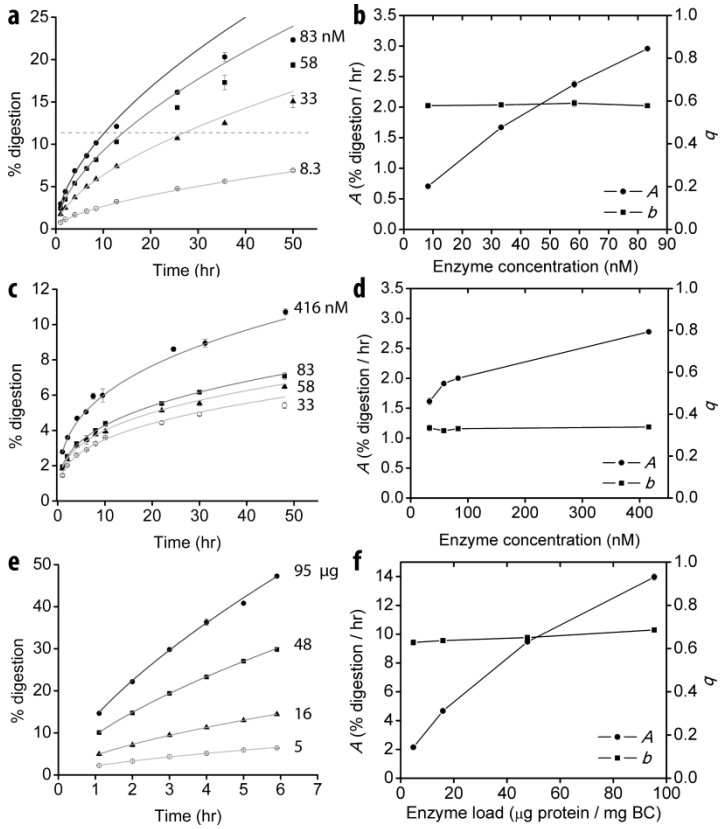
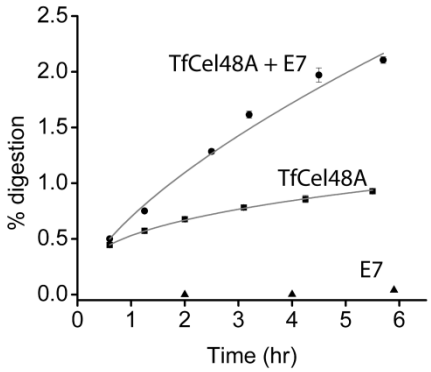


Figure 2.3: Dependence of *A* and *b* on enzyme concentration. Different concentrations of TfCel9A ((a) and (b)) TfCel5A ((c) and (d)), and *T. fusca* crude supernatant ((e) and (f)) on BC. All time course data were obtained in triplicate. Error bars in (a), (c), and (e) represent standard deviation. Time course data were fit with Eqn. 2.6 and the obtained *A* and *b* values are plotted in (b), (d), and (f); error bars represent standard error. Dashed line in (b) indicates the approximate "drop-off" value, above which digestion rate decreases more rapidly than predicted by the initial fit.

parameters that quantify the different factors involved in digestion (e.g. enzyme binding, change in the average degree of polymerization, change in cellulose crystallinity, etc.).

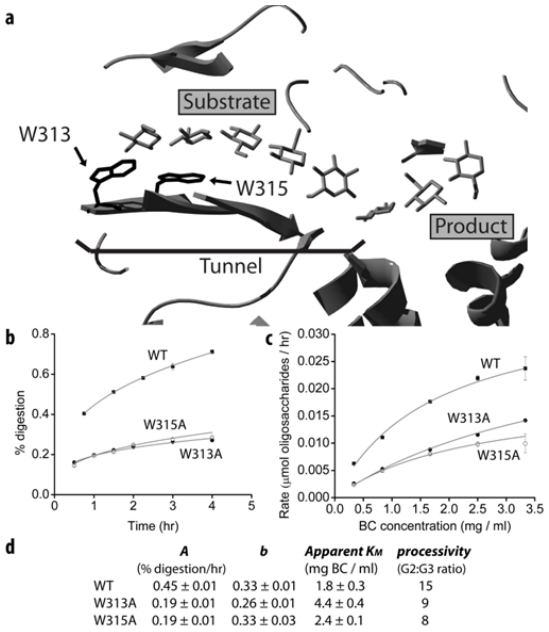
Proposed here is an alternative approach, which treats cellulose digestion similarly to classical kinetics, but replaces the specific activity constant with a coefficient that is dependent on time (extent of substrate digestion). The results indicate that cellulases have different time-dependencies of their activity coefficients when acting on insoluble substrates, as quantified by parameter *b* (curvature of the time course profile) of Eqns. 4 and 6. This parameter is not dependent on the enzyme concentration within the tested range (Fig. 2.3) and appears to be only slightly dependent on temperature (Fig. 2.2). These observations support the idea that the value



	A (% digestion/hr)	*b*
TfCel48A	0.53 ± 0.01	0.33 ± 0.01
TfCel48A + E7	0.70 ± 0.01	0.65 ± 0.02

Figure 2.4: Effect of the auxiliary protein *T. fusca* E7 (AA10) on the digestion of BC by a weak *T. fusca* exocellulase TfCel48A. Time course data were collected in triplicate and were fit to Eqn. 2.6. The calculated parameters are presented in the Table inset. E7 alone produces only trace amounts of soluble sugars and its time course profile does not follow Eqn. 2.6.

of *b* quantifies the intrinsic ability of the cellulase to digest bulk cellulose. Based on the derivation of Eqn. 2.4, the value of *b* will fall between zero, when no products are produced over time, and one, when the specific activity is constant, as is in classical kinetics. A relatively robust cellulase, which can effectively hydrolyze recalcitrant portions of cellulose, will have a higher *b* value than a cellulase that is only effective on the easily accessible fractions of the substrate. This holds true for *T. fusca* cellulases acting on BC (Table 2.2). In particular, TfCel9A, which is the most effective *T. fusca* cellulase on crystalline substrates (*4*) has the highest *b* value of 0.58 ± 0.02. Other cellulases, which are variably effective in crystalline cellulose digestion, have *b* values in the range from 0.31 to 0.50. The *b* value does not change when the CBM is removed from the cellulase (Table 2.2), which leads to a reduced binding efficiency of the cellulase (*5*, *24*). This suggests that the CBM of the tested enzymes (TfCel9A and TfCel48A) does not directly assist the CD in substrate digestion, as has been suggested for some cellulases in the literature (*25*-*27*). The constant *b* parameter value with and without the CBM attached to the CD also indicates that the time dependence of the cellulase's specific activity is not related to surface diffusion limitations, which form the basis of fractal-like kinetic modeling (*2*). Rather, in the case of cellulose digestion, the continuously decreasing specific activity is most likely due to



d

	A (% digestion/hr)	b	Apparent K_M (mg BC / ml)	processivity (G2:G3 ratio)
WT	0.45 ± 0.01	0.33 ± 0.01	1.8 ± 0.3	15
W313A	0.19 ± 0.01	0.26 ± 0.01	4.4 ± 0.4	9
W315A	0.19 ± 0.01	0.33 ± 0.03	2.4 ± 0.1	8

Figure 2.5: Role of W313 and W315 in the digestion of BC by TfCel48A. (a) The tunnel and active site of TfCel48A. W313 and W315 residues are located at the mouth and inside the tunnel respectively; (b) time course profiles of WT, W313A, and W315A on BC (N=3), fitted with Eqn. 6; (c) apparent K_M experiments with WT, W313A, and W315A on BC; (d) Eqn. 6 parameter values, processivity, and apparent K_M of WT and mutated TfCel48A. Processivity is calculated as a ratio of cellobiose (G2) to cellotriose (G3).

increasing substrate recalcitrance as easier to digest fractions are preferentially hydrolyzed by the enzyme(s).

The value of *A* follows the Arrhenius relationship as a function of temperature (Fig. 2.2), which is consistent with its proposed significance. Further support is provided by the data presented in Table 2.2. When the CBM is removed from TfCel9A and TfCel48A, the *A* value is significantly reduced, presumably due to the lower binding efficiency of the CD alone. Both exocellulases (TfCel48A and TfCel6b) have much lower *A* parameters than the endocellulases, which probably reflects the fewer productive binding sites (chain ends) available to these enzymes. Finally, the *A* value is strongly dependent on enzyme concentration (Fig. 2.3), as would be expected for the productively bound concentration at low enzyme-substrate ratios. The fact that no two *T. fusca* cellulases have the same combination of *A* and *b* values suggests that no two cellulases act in the same way. It would be of particular interest to identify the unique

features of the different endocellulases (Cel5A, Cel6B, and Cel9B in *T. fusca*), as they may have unique roles in the synergistic mixtures during substrate digestion.

A potentially useful aspect of the model is the variable range of the extent of digestion over which the model is applicable. At least some individual cellulases appear to reach a "drop-off" value, above which the rate of digestion decreases more rapidly than would be predicted by the *b* parameter. For TfCel9A (Fig. 2.3a) and TfCel48A (Fig. 6.1) acting on BC the "drop-off" values are approximately at 11% and 2% digestion respectively. On the other hand, the model fits *T. fusca* and *T. reesei* crude supernatant digestion of BC up to at least 50% (Fig. 2.1 c and d). This observation suggests that as different biomass degrading enzymes are combined in synergistic mixtures, their ability to maintain initial-like conditions are enhanced. It is likely that during cellulose digestion by individual cellulases or incomplete synergistic mixtures, factors other than increased substrate recalcitrance contribute to the drop in the digestion rate earlier, which results in a more significant rate decline than would be predicted from the initial data. Such factors may be attributed, for example, to an increasing fraction of unproductively bound enzymes (*28, 29*) or the formation of obstacles on the eroded surface of the substrate (*30*). A constant *b* value over a large extent of digestion by the crude extract of *T. fusca* and *T. reesei* suggests that relief of such factors plays a significant role in the synergism between cellulases with complementary properties (*30, 31*).

Eqns. 2.4 and 2.6 apply to both individual cellulases and their mixtures (Fig. 2.1). This is an important feature of the model, as it makes it possible to design experiments aimed at understanding synergistic interactions between cellulases and other important proteins. In comparison to the individual cellulases, the *T. fusca* crude supernatant time course profile on BC has the highest *b* value, which suggests that the synergistic interactions between proteins in the

crude supernatant enhance the ability of the mixture to digest a crystalline substrate more effectively (not only faster) than any cellulase alone. Cellulases can be combined in mixtures of two or more and the changes in the parameter values can help determine the mechanisms of synergistic interactions between those enzymes. The data presented for BC digestion by TfCel48A (Fig. 2.4) with and without the disruptive *T. fusca* AA10 protein E7 shows that E7 increases the *b* and *A* parameter values of Cel48A by 97% and 32% respectively. This is consistent with other studies, which show that AA10 (CBM33) proteins are likely to disrupt the recalcitrant portions of crystalline cellulose with minimal formation of soluble products (*15, 32*). The relatively minor increase in the *A* parameter indicates that the specific activity of TfCel48A acting on BC probably is not increased in the presence of E7. Rather, E7 seems to increase the accessible fraction of the substrate, which results in a higher fraction of productively bound enzyme. Our group is also carrying out similar time course experiments as part of a study aimed to understand the synergistic interactions between TfCel9A and TfCel48A.

It is important to verify that Eqn. 2.6 applies to biomass-degrading enzymes from other organisms. For this reason the crude supernatant of *T. reesei* was tested on BC and Avicel®. As Fig. 2.1d shows, Eqn. 2.6 fits well *T. reesei* crude supernatant digestion of BC and Avicel® up to at least 50% and 20% digestion respectively. While comparison of the *T. fusca* and *T. reesei* crude supernatant time course profiles may provide interesting information about the possible differences between the bacterial and fungal biomass digestion mechanisms, such a study would require additional rigorous experiments. For this reason, no conclusions are made based on the experiments described here, but the results demonstrate that the kinetic model described by Eqns. 4 and 6 is applicable to both bacterial and fungal systems.

Dependence of the cellulose hydrolysis rate coefficient on time (extent of digestion) had been previously shown by Valjamae and co-workers (*11*). Similarly, Ohmine and co-workers demonstrated that the time course profiles of cellulose digestion can be defined by two empirically determined parameters that quantify the interaction between a cellulase mixture and its substrate (*33*). In both studies, however, only mixtures of cellulases were tested under pseudo-first order (with respect to the substrate) conditions (i.e. much higher enzyme to substrate ratios). This leads to important differences between the significance of our model and those published in the above studies. All three models contain a parameter that quantifies the shape of the time course profile (*b* parameter in the proposed model). In both of the above studies this parameter is dependent on enzyme-substrate ratio, while the *b* parameter in this model is constant within the tested range for individual cellulases. On the other hand, a slight dependence of the *b* value on enzyme load was observed for *T. fusca* crude digestion of BC. This is consistent with the explanation offered by Valjamae et al. (*11*) that at increasing enzyme loads the distances between the synergistically acting enzymes decrease, leading to a more efficient digestion of the substrate. Since there is no auto-synergism by individual cellulases, an increase in enzyme to substrate ratio, should not alter the shape of the digestion profile for a single cellulase. In addition, by working at much lower enzyme to substrate ratios than those used in the above studies, we avoid the contribution of substrate depletion to the multiple factors responsible for the continuous drop in the digestion rate of cellulose.

In comparison to the more complex kinetic models of enzymatic cellulose digestion developed to date, our model has unique advantages for the studies of mechanisms by which cellulases degrade insoluble cellulose. Only small amounts of enzyme are required for initial studies. No preliminary information about the cellulase is required; the empirically obtained *A*

and *b* parameters are themselves meaningful measures of the cellulase properties and the *b* parameter in particular can be used to compare cellulases across different studies. Correlation between induced modifications to the cellulases (e.g. TfCel48A mutations, Fig. 2.5) or to their substrates (e.g. PASC derived from Avicel®) and the corresponding changes in parameter values can provide information about the effects of the induced modifications. Because the model applies both to individual cellulases and their mixtures, one can design synergistic experiments that lead to a better understanding of the specific interactions between the enzymes of interest (e.g. TfCel48a in combination with E7, Fig. 2.4). Hence, the presented model can be useful for basic studies of enzymatic digestion of cellulose, as well as for the design of improved synergistic biomass-degrading cocktails and pretreatment methods of lignocellulosic biomass.

References

1. Michaelis, L., and Menten, M. L. (1913) The kenetics of the inversion effect., *Biochem Z 49*, 333-369.
2. Kopelman, R. (1988) Fractal Reaction-Kinetics, *Science 241*, 1620-1626.
3. Zhang, Y. H., Cui, J., Lynd, L. R., and Kuang, L. R. (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure, *Biomacromolecules 7*, 644-648.
4. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects, *Biotechnol Bioeng 42*, 1002-1013.
5. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of Thermomonospora fusca E4 in cellulose hydrolysis, *J Bacteriol 180*, 1709-1714.

6. Irwin, D., Leathers, T. D., Greene, R. V., and Wilson, D. B. (2003) Corn fiber hydrolysis by Thermobifida fusca extracellular enzymes, *Appl Microbiol Biotechnol 61*, 352-358.

7. Walker, L. P., Wilson, D. B., Irwin, D. C., Mcquire, C., and Price, M. (1992) Fragmentation of Cellulose by the Major Thermomonospora-Fusca Cellulases, Trichoderma-Reesei Cbhi, and Their Mixtures, *Biotechnol Bioeng 40*, 1019-1026.

8. Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sorlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides, *Science 330*, 219-222.

9. Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K. C., Borch, K., and Westh, P. (2011) A kinetic model for the burst phase of processive cellulases, *Febs J 278*, 1547-1560.

10. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., and Westh, P. (2012) Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A, *J Biol Chem 287*, 18451-18458.

11. Valjamae, P., Kipper, K., Pettersson, G., and Johansson, G. (2003) Synergistic cellulose hydrolysis can be described in terms of fractal-like kinetics, *Biotechnol Bioeng 84*, 254-257.

12. Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S., and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family, *Biochemistry 49*, 3305-3316.

13. Phillips, C. M., Beeson, W. T., Cate, J. H., and Marletta, M. A. (2011) Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by Neurospora crassa, *ACS Chem Biol*.

14. Langston, J. A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E., and Sweeney, M. D. (2011) Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61, *Appl Environ Microbiol*.

15. Moser, F., Irwin, D., Chen, S., and Wilson, D. B. (2008) Regulation and characterization of Thermobifida fusca carbohydrate-binding module proteins E7 and E8, *Biotechnol Bioeng 100*, 1066-1077.

16. Zhou, W., Irwin, D. C., Escovar-Kousen, J., and Wilson, D. B. (2004) Kinetic studies of Thermobifida fusca Cel9A active site mutant enzymes, *Biochemistry 43*, 9655-9663.

17. Vuong, T. V., and Wilson, D. B. (2009) Processivity, synergism, and substrate specificity of Thermobifida fusca Cel6B, *Appl Environ Microbiol 75*, 6655-6661.

18. Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2011) Determination of the molecular states of the processive endocellulase Thermobifida fusca Cel9A during crystalline cellulose depolymerization, *Biotechnol Bioeng*.

19. Fox, J. M., Levine, S. E., Clark, D. S., and Blanch, H. W. (2012) Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes, *Biochemistry 51*, 442-452.

20. Reverbel-Leroy, C., Parsiegla, G., Moreau, V., Juy, M., Tardif, C., Driguez, H., Belaich, J. P., and Haser, R. (1998) Crystallization of the catalytic domain of Clostridium cellulolyticum CeLF cellulase in the presence of a newly synthesized cellulase inhibitor, *Acta Crystallogr D Biol Crystallogr 54*, 114-118.

21. Koivula, A., Kinnari, T., Harjunpaa, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T. A., and Teeri, T. T. (1998) Tryptophan 272: an essential determinant of crystalline cellulose degradation by Trichoderma reesei cellobiohydrolase Cel6A, *FEBS Lett 429*, 341-346.

22. Horn, S. J., Sikorski, P., Cederkvist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M., and Eijsink, V. G. (2006) Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides, *Proc Natl Acad Sci U S A 103*, 18089-18094.

23. Zhang, S., and Wilson, D. B. (1997) Surface residue mutations which change the substrate specificity of Thermomonospora fusca endoglucanase E2, *J Biotechnol 57*, 101-113.

24. Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca, *Eur J Biochem 267*, 4988-4997.

25. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, A. J., and Kilburn, D. G. (1991) Non-Hydrolytic Disruption of Cellulose Fibers by the Binding Domain of a Bacterial Cellulase, *Bio-Technol 9*, 1096-1099.

26. Teeri, T., Reinikainen, T., Ruohonen, L., Jones, T. A., and Knowles, J. K. C. (1992) Domain Function in Trichoderma-Reesei Cellobiohydrolases, *J Biotechnol 24*, 169-176.

27. Esteghlalian, A. R., Srivastava, V., Gilkes, N. R., Kilburn, D. G., Warren, R. A., and Saddle, J. N. (2001) Do cellulose binding domains increase substrate accessibility?, *Appl Biochem Biotechnol 91-93*, 575-592.

28. Yang, B., Willies, D. M., and Wyman, C. E. (2006) Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion, *Biotechnol Bioeng 94*, 1122-1128.

29. Eriksson, T., Karlsson, J., and Tjerneld, F. (2002) A Model Explaining Declining Rate in Hydrolysis of Lignocellulose Substrates with Cellobiohydrolase I (Cel7A) and Endoglucanase I (Cel7B) of Trichoderma reesei, *Appl Biochem Biotechnol 101*, 41-60.

30. Valjamae, P., Sild, V., Nutt, A., Pettersson, G., and Johansson, G. (1999) Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I, *Eur J Biochem 266*, 327-334.

31. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttila, M., Ando, T., and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface, *Science 333*, 1279-1282.

32. Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunaes, A. C., Stenstrom, Y., Mackenzie, A., Sorlie, M., Horn, S. J., and Eijsink, V. G. (2011) Cleavage of cellulose by a CBM33 protein, *Protein Sci 20*, 1479-1483.

33. Ohmine, K., Ooshima, H., and Harano, Y. (1983) Kinetic study on enzymatic hydrolysis of cellulose by cellulose from Trichoderma viride, *Biotechnol Bioeng 25*, 2041-2053.

CHAPTER 3

Determination of the molecular states of the processive endocellulase *Thermobifida fusca* TfCel9A during crystalline cellulose depolymerization.

Abstract:

Detailed understanding of cell wall degrading enzymes is important for their modeling and industrial applications, including in the production of biofuels. Here we used TfCel9A, a processive endocellulase from *Thermobifida fusca*, to demonstrate that cellulases that contain a catalytic domain (CD) attached to a cellulose binding module (CBM) by a flexible linker exist in three distinct molecular states. By measuring the ability of a soluble competitor to reduce TfCel9A activity on an insoluble substrate, we show that the most common state of TfCel9A is bound via its CBM, but with its CD unoccupied by the insoluble substrate. These findings are relevant for kinetic modeling and microscopy studies of modular glycoside hydrolases.

Introduction

Kinetic modeling of cellulase activity on crystalline cellulose is complicated by the presence of two distinct phases: the heterogeneous insoluble substrate and short oligosaccharides in solution. Digestion of cellulose requires a number of distinct steps. In order to initiate cleavage of its substrate, the cellulase must first bind to the bulk cellulose and capture an individual cellulose chain in its active site. Once the chain is positioned appropriately in the active site, the β-1,4 glycosydic bond is hydrolyzed and the cellulase either releases the chain or processively translates along the chain carrying out additional cleavages. For the purpose of modeling, it has been suggested that binding and initiation can be treated as a one (Okazaki and Moo-Young 1978; Zhang and Lynd 2006; Zhou et al. 2009a; Zhou et al. 2009b) or a two (Levine et al. 2010) step process. Here, we present experimental evidence for TfCel9A that binding and cleavage of the insoluble substrate are two separate processes and should, therefore, be treated as two distinct steps in cellulose depolymerization. Our data also suggest that a large fraction of

TfCel9A adsorbed to the crystalline substrate is bound to the surface via its CBM2, but with an unoccupied CD, whose active site can be accessed by a soluble competitor present in solution.

TfCel9A fluorescent labeling was done by Jose Moran-Mirabal in Larry Walker's laboratory in Cornell University.

Methods and Materials

Enzyme production, purification, and labeling: Native TfCel9A and TfCel9A-68 were expressed and purified as previously described (Li et al. 2007; Zhang et al. 1995). Native TfCel9A was fluorescently labeled previously, as described in (Moran-Mirabal et al. 2009). It was verified that the activity of labeled TfCel9A was the same as that of the native enzyme (Figure 3.1).

Cellulose preparation: A stock solution of bacterial cellulose (BC, Monstanto Cellulon, Monsanto, San Diego, CA) was prepared as previously described (Jeoh et al. 2002).

Hydrolysis and binding assays: Hydrolysis reactions were prepared in Eppendorf Protein LoBind 2 ml tubes (Eppendorf, Hamburg, Germany). 0.5 mg BC was mixed with either native TfCel9A or TfCel9A-68 in a total of 600 µl 50 mM sodium acetate buffer, pH 5.5. Different amounts of the two enzymes were used in order to achieve similar extent of digestion. In the endpoint measurements without the addition of G4 3% of the substrate was digested after 8 hour incubation with 5 pmoles of the native enzyme and 2.5% of the substrate was digested after four hour incubation with 15 pmoles of TfCel9A-68. When applicable, G4 was added to the reactions together with the enzyme. Reactions were performed in triplicates and were incubated for the indicated amount of time at 50 °C. To stop the reaction, samples were centrifuged at 10,000 g for 5 minutes at 4 °C. For binding measurements 150 µl of the supernatant was transferred to a new tube and frozen at -20 °C.

Extent of hydrolysis was determined by measuring soluble oligosaccharide concentrations present in the reaction buffer after incubation. Soluble oligosaccharides were measured by refractive index using a Shimadzu HPLC system (Shimadzu, Columbia, MD, USA). Hydrolysis samples (50 µl injection volume) were eluted with water at 0.600 ml/min through an HPX-87P Bio-Rad Aminex column. Oligosaccharide peaks were fitted to Gaussian functions using OriginPro8 (Origin Lab, Northhampton, MA, USA) and concentrations were determined using standards of known concentrations. For accurate comparison total soluble sugars were converted to µmoles glucose. When G4 was added to the reactions, its exact concentration was determined from control reactions that contained no enzyme. The measured amount of added G4 per reaction was then subtracted from the total soluble sugar in the reaction buffer, allowing for accurate measurements of the soluble sugars released from the depolymerization of BC even at short incubation times.

Native TfCel9A binding was determined as follows. All fluorescence measurements were done using Microfluor® 2 Black flat bottom 96 well plates (Thermo Electron Corp., Milford, MA, USA) and a Biotek Synergy 4 plate reader (Biotek Instruments, Winoosli, VT, USA), with the excitation wavelength set at 485 nm and the emission wavelength set at 528 nm. A standard curve of fluorescence as a function of enzyme concentration was prepared using the enzyme master solution used for the activity assays. The first set of wells contained the same concentration of enzyme that was added to the hydrolysis assays. From these wells the solution was serially diluted by a factor of 2 across the plate. The fluorescence intensity measured in the first set of wells was set as 100 % enzyme added to the reactions. Fluorescence intensity measured from the hydrolysis assays was converted to percent enzyme present in the reaction buffer using the standard curve. Percent bound enzyme was calculated as the difference between

the total enzyme added to the reaction and the amount of enzyme present in the reaction buffer at the completion of the assay.

Results and Discussion

In the following experiments we used two versions of TfCel9A: native enzyme and a mutant without CBM2 (Cel9A-68). The native enzyme was labeled with a fluorescent dye (Alexa Fluor 488, degree of labeling of 1.5), which allowed for accurate measurements of TfCel9A concentrations down to the 0.1 nM range and made it possible to accurately determine enzyme binding directly in the activity assays. Activity assays were carried out by incubating enzyme with bacterial cellulose (BC) in 50 mM sodium acetate buffer, pH 5.5 at 50 °C. When applicable, G4 was added to the reactions at the same time as the enzyme. Soluble oligosaccharides produced during substrate digestion were analyzed via high performance liquid chromatography (HPLC). The percentage of native bound enzyme was determined by measuring fluorescence of the reaction buffer as part of the activity assays and comparing it with reference (no substrate) controls.

In order to determine whether the binding of TfCel9A via CBM2 automatically resulted in the binding of its CD, we measured endpoint activity of TfCel9A on BC in the presence of increasing concentrations of G4 (Figure 3.2A), its primary product (Irwin et al. 1998). Once in solution, G4 is rapidly cleaved by TfCel9A to cellotriose, cellobiose, and glucose, and is therefore a competitive substrate that can reduce TfCel9A activity on insoluble substrates. It is worth noting that preliminary experiments with cellotriose and cellobiose showed that these oligosaccharides do not reduce activity of TfCel9A on BC at the same concentrations that were tested for G4 (Figure 3.3), which is consistent with their lower binding affinities in the active site (Zhou et al. 2004). Similarly, glucose does not inhibit any of the *T. fusca* cellulases at the

concentration normally observed in various activity assays (data not shown), presumably due to its low binding affinity. In the conditions used for the activity assays more than 90% of the native TfCel9A bound to BC (Figure 3.2B). This is about 10% higher binding than previously published data, which was obtained at room temperature using significantly higher enzyme loads (Irwin et al. 1998). We found that increasing concentrations of G4 lead to a substantial decrease in the total sugar produced by TfCel9A. Importantly, the presence of increasing concentrations of G4 had only a negligible effect on the percentage of TfCel9A bound to BC, which cannot account for the observed drop in activity. Hence, the data indicated that although the majority of TfCel9A was bound to the cellulose surface, its CD was accessible to G4, a soluble competitor. Since TfCel9A is known to bind primarily via its CBM2 (Irwin et al. 1998), inhibition by G4 indicated that a significant fraction of the enzyme was bound to BC via CBM2, with an unbound, unoccupied CD.

In order to further verify our conclusion, we also looked at the effect of increasing G4 concentrations on TfCel9A-68, which does not have CBM2 (Figure 3.2A). Previously published binding assays, carried out at room temperature, showed that TfCel9A-68 does not measurably bind to BC (Irwin et al. 1998). Low levels of binding (up to 15%) could be attained by performing the same binding assays at 4 °C in order to significantly lower the activity of the enzyme (Li et al. 2007; Zhou et al. 2004). Binding of TfCel9A-68 was also induced by introducing mutations to the catalytic acid and base residues in the active site (Zhou et al. 2004). We found that G4 inhibited the hydrolytic activity of TfCel9A-68 to the same extent as that of the native enzyme, indicating that the CD of TfCel9A is equally accessible by a soluble substrate whether or not it is bound to the surface of crystalline cellulose via its CBM2. Thus, the G4

inhibition data obtained for TfCel9A-68 supported our conclusion drawn from the G4 inhibition experiments using the native TfCel9A.

Since the data shown in Figure 3.2 reflect the endpoint measurement of activity and binding after several hours of incubation, it was important to verify that the binding of TfCel9A to BC is not affected by the presence of G4 in the initial stages of the digestion process. One possible scenario, which could not be ruled out from the data shown in Figure 3.2, was that G4 reduced the rate of TfCel9A binding to BC, which would predominantly affect TfCel9A activity in the initial stages of digestion. In order to determine whether high concentration of G4 affected the binding rate of TfCel9A to BC we measured the activity and binding of TfCel9A over time in the presence and absence of 85 μM G4 (Figure 3.4). Inhibition of TfCel9A activity on BC by G4 was already evident after 1 hour incubation time, while its binding was unaffected. Therefore, the reduced activity of TfCel9A in the presence of G4 cannot be attributed to a lower binding rate of the enzyme even at the early stages of digestion.

The ability of G4 to inhibit TfCel9A activity on BC without a measurable effect on the binding of TfCel9A indicates that the active site of the CD is frequently available even when the enzyme is bound to the surface of BC. Most binding events are presumed to occur via CBM2, as its removal leads to significantly reduced binding of TfCel9A to crystalline substrate (Irwin et al. 1998). The similar level of inhibition by G4 observed for TfCel9A and TfCel9A-68 (no CBM2) suggests that at any given instant the majority of the bound enzyme contains an unbound CD, allowing for a soluble substrate to bind in the active site. Hence, we can infer from our observations that in the presence of crystalline substrate, native TfCel9A can be found in three distinct molecular states shown in Figure 3.5: i) free in solution, with an unoccupied CD and an unbound CBM2; ii) bound to the cellulose surface via CBM2, with an unoccupied CD, which

can bind and cleave soluble substrate; iii) bound to the surface via CBM2 and CBM3c with the active site of the CD occupied by a single cellulose chain undergoing hydrolysis.

Clear experimental evidence of the three unique states of TfCel9A imply that the binding of the enzyme via CBM2 and the initiation of cleavage are two distinct events, which should be treated as such in the kinetic modeling of cellulases that contain a CBM separated from the CD by a flexible linker. Determining the exact fraction of the inactively bound TfCel9A is complicated by a number of factors, including the extent of substrate digestion, and is beyond the scope of this report. However, our results suggest that the most common state for TfCel9A is bound via its CBM2 with an unoccupied CD. This conclusion is supported by the similar level of inhibition by G4 of both TfCel9A and TfCel9A-68 despite the fact that TfCel9A-68 does not measurably bind to crystalline cellulose and the high level binding of TfCel9A is unaffected by the high concentrations of G4 (Figures 3.2B and 3.4B). This is an important consideration for various approaches to studying TfCel9A and related cellulases. In particular, high resolution, single molecule microscopy studies can further reveal the differences between the unproductively and productively bound enzymes with regard to their molecular displacements and reversibility of binding. It would also be useful to determine what fraction of unproductively bound enzymes is bound irreversibly and whether the ratio of inactively bound to active enzymes increases with substrate digestion. Such information may prove useful in explaining the synergistic behavior between cellulases and in determining the limiting steps in the degradation of crystalline cellulose. One way synergistic partners enhance each other's activity, for example, may be by increasing the ratio of the productively bound states, especially toward the later stages of digestion (Valjamae et al. 1999).

In summary, we have shown that soluble G4 is a substrate that competes with insoluble cellulose for the catalytic site of *Tf*Cel9A. We compared the level of G4 inhibition of TfCel9A with and without its CBM2, which is responsible for binding the enzyme to the substrate. The level of inhibition was similar for the two versions of TfCel9A, indicating that even when the enzyme is bound via its CBM2, its CD is usually unoccupied. We thus conclude that the most common state of the native TfCel9A in crystalline substrate digestion is bound on the substrate surface via its CBM2 with an unoccupied CD. The experiments presented here can serve as a general method to determine the relative abundances of the three states – unbound, unproductively bound, and productively bound – for other CBM-containing cellulases. This information can be used to develop better models of cellulase activities, compare different cellulases to each other, and better understand cellulase behavior in synergistic mixtures.

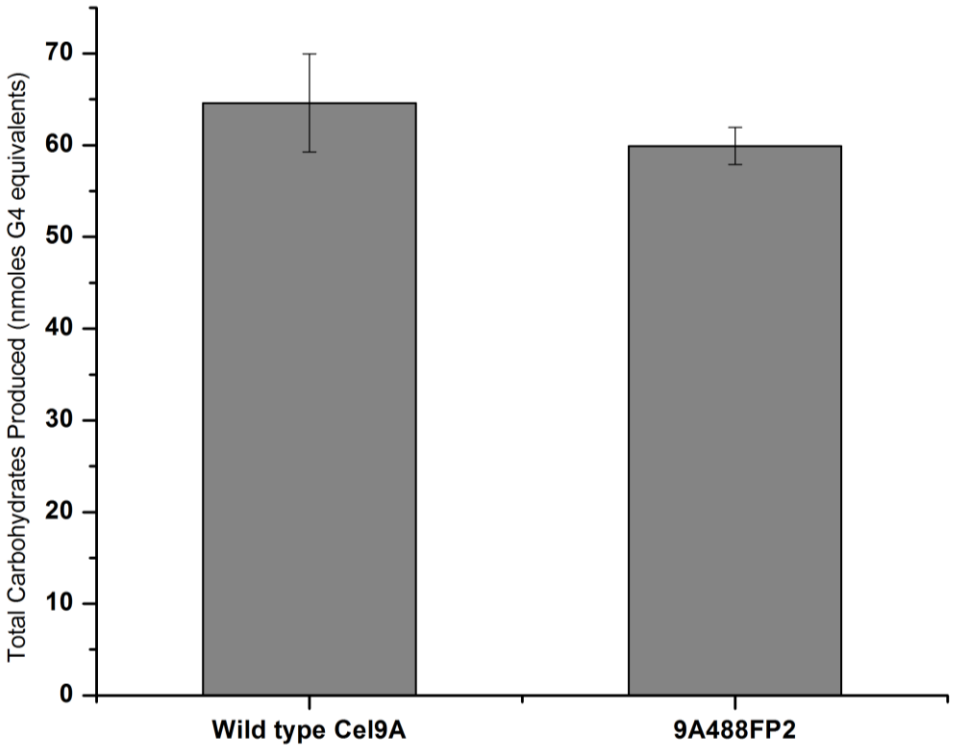


Figure 3.1: Activity of wild type and labeled TfCel9A on BC, showing that labeling does not compromise the activity of TfCel9A

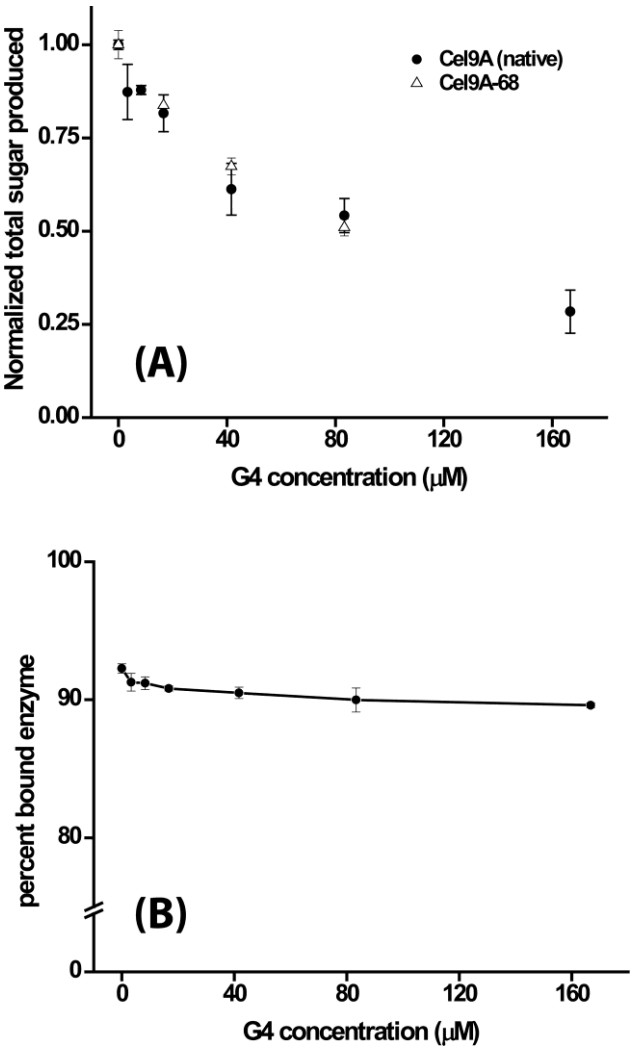


Figure 3.2: (A) Normalized activity of TfCel9A and TfCel9A-68 on BC in the presence of increasing concentration of G4. Reactions with TfCel9A contained 5 pmoles enzyme and were incubated for 8 hours. Reactions with TfCel9A-68 contained 15 pmoles enzyme and were incubated for 4 hours. Similar % digestion was attained in both cases (3 and 2.5 % for Cel9A and Cel9A-68 respectively without the addition of G4). (B) Binding of TfCel9A in the presence of G4. Binding measurements were taken directly from the activity assays.

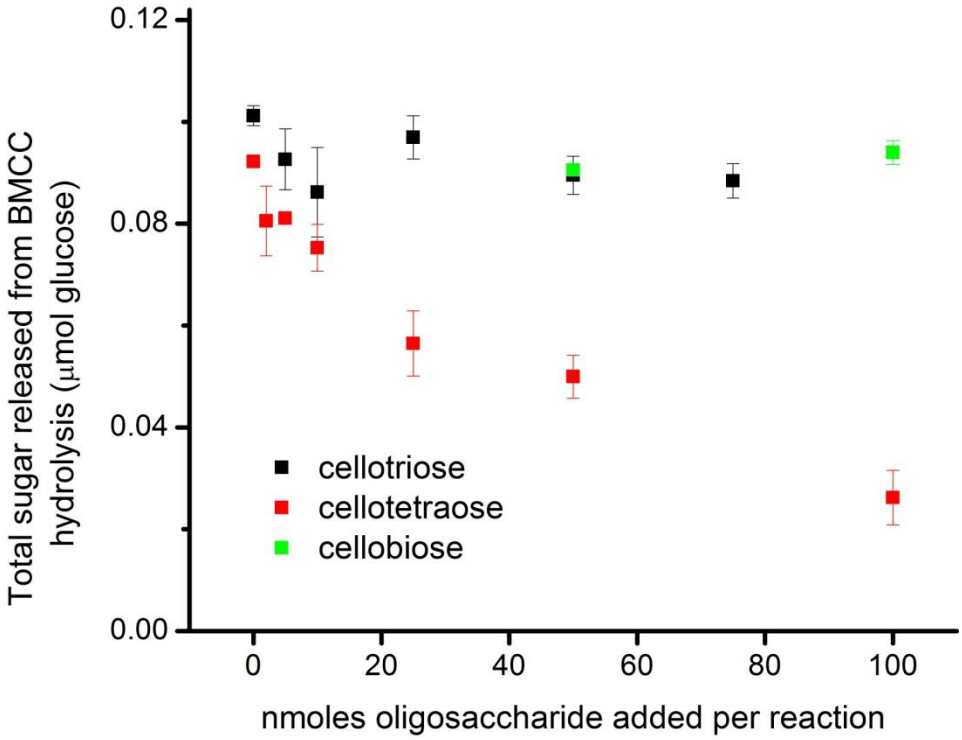


Figure 3.3: Effect of soluble oligosaccharide addition on the activity of native TfCel9A on BC. Only the addition of G4 (cellotetraose) resulted in the reduction of BC hydrolysis.

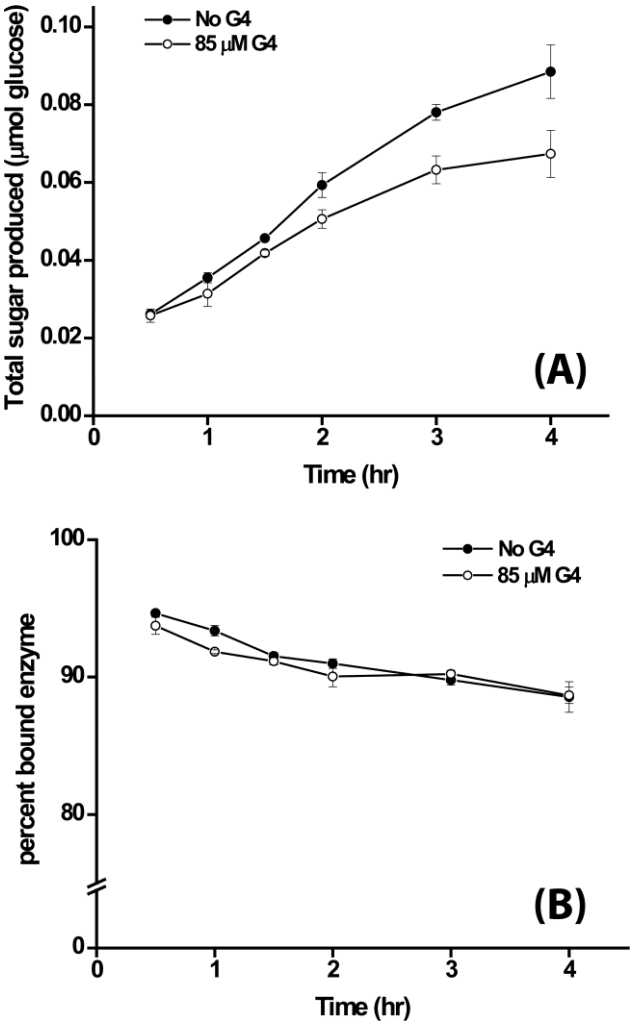


Figure 3.4: (A) Activity of TfCel9A on BC over time in the presence and absence of 85 μM G4. Each reaction contained 5 pmoles enzyme. (B) Binding of TfCel9A in the same set of reactions.

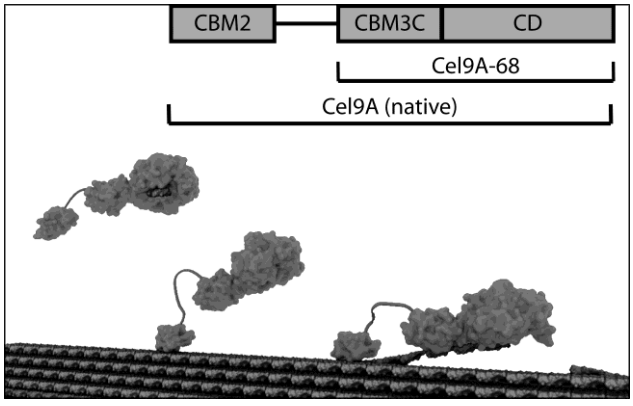


Figure 3.5: The three states of TfCel9A in the presence of crystalline substrate. The most common state is bound via CBM2 with an unbound, unoccupied CD. Inset shows the modular structure of TfCel9A.

References

Irwin D, Shin DH, Zhang S, Barr BK, Sakon J, Karplus PA, Wilson DB. 1998. Roles of the catalytic domain and two cellulose binding domains of Thermomonospora fusca E4 in cellulose hydrolysis. Journal of bacteriology 180(7):1709-14.

Jeoh T, Wilson DB, Walker LP. 2002. Cooperative and competitive binding in synergistic mixtures of Thermobifida fusca cellulases Cel5A, Cel6B, and Cel9A. Biotechnology progress 18(4):760-9.

Levine SE, Fox JM, Blanch HW, Clark DS. 2010. A mechanistic model of the enzymatic hydrolysis of cellulose. Biotechnology and bioengineering 107(1):37-51.

Li Y, Irwin DC, Wilson DB. 2007. Processivity, substrate binding, and mechanism of cellulose hydrolysis by Thermobifida fusca Cel9A. Applied and environmental microbiology 73(10):3165-72.

Merino ST, Cherry J. 2007. Progress and challenges in enzyme development for biomass utilization. Advances in biochemical engineering/biotechnology 108:95-120.

Moran-Mirabal JM, Corgie SC, Bolewski JC, Smith HM, Cipriany BR, Craighead HG, Walker LP. 2009. Labeling and purification of cellulose-binding proteins for high resolution fluorescence applications. Analytical chemistry 81(19):7981-7.

Okazaki M, Moo-Young M. 1978. Kinetics of enzymatic hydrolysis of cellulose: analytical description of a mechanistic model. Biotechnology and bioengineering 20(5):637-63.

Oliveira OV, Freitas LCG, Straatsma TP, Lins RD. 2009. Interaction between the CBM of Cel9A from Thermobifida fusca and cellulose fibers. Journal of Molecular Recognition 22(1):38-45.

Perlack RD, Wright LL, Turhollow AF, Graham R, L., Stokes B, J., Erbach DC. 2005. Biomass as Feedstock for a Bioenergy and Bioproducts Industry: the Technical Feasibility of a Billion-ton Annual Supply. Oak Ridge: Oak Ridge National Laboratory.

Sakon J, Irwin D, Wilson DB, Karplus PA. 1997. Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. Nature structural biology 4(10):810-8.

Shoseyov O, Shani Z, Levy I. 2006. Carbohydrate binding modules: biochemical properties and novel applications. Microbiology and molecular biology reviews : MMBR 70(2):283-95.

Valjamae P, Sild V, Nutt A, Pettersson G, Johansson G. 1999. Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. European journal of biochemistry / FEBS 266(2):327-34.

Wen F, Nair NU, Zhao H. 2009. Protein engineering in designing tailored enzymes and microorganisms for biofuels production. Current opinion in biotechnology 20(4):412-9.

Zhang S, Lao G, Wilson DB. 1995. Characterization of a Thermomonospora fusca exocellulase. Biochemistry 34(10):3386-95.

Zhang YH, Lynd LR. 2006. A functionally based model for hydrolysis of cellulose by fungal cellulase. Biotechnology and bioengineering 94(5):888-98.

Zhou W, Hao Z, Xu Y, Schuttler HB. 2009a. Cellulose hydrolysis in evolving substrate morphologies II: Numerical results and analysis. Biotechnology and bioengineering 104(2):275-89.

Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB. 2004. Kinetic studies of Thermobifida fusca Cel9A active site mutant enzymes. Biochemistry 43(30):9655-63.

Zhou W, Schuttler HB, Hao Z, Xu Y. 2009b. Cellulose hydrolysis in evolving substrate morphologies I: A general modeling formalism. Biotechnology and bioengineering 104(2):261-74.

CHAPTER 4

Determination of the catalytic base in family 48 glycoside hydrolases**.**

Abstract

The catalytic base in family 48 glycoside hydrolases has not been previously established experimentally. Based on structural and modeling data published to date, we used site directed mutagenesis and azide rescue activity assays to show definitively that the catalytic base in *Thermobifida fusca* TfCel48A is aspartic acid 224. Of the tested mutants only TfCel48A D224E retained partial activity on soluble and insoluble substrate. In azide rescue experiments only D224G, the smallest residue tested, showed an increase in activity with added azide.

Introduction

Family 48 cellulases form an important part of the bacterial cellulose degrading system. Despite their low specific activity on crystalline cellulose, they have high synergistic activity with other cellulases (*1-3*). The three most studied members of the family are Cel48F from *Clostridium cellulolyticum*, Cel48S from *Clostridium thermocellum*, and TfCel48A from *Thermobifida fusca*. Cel48F and Cel48S are major components in the cellulose-degrading multienzyme complexes called cellulosomes (*4-6*), while TfCel48A forms 34% of the total extracellular cellulases produced by *T. fusca* (*7*). The catalytic domain structures of Cel48F (*6, 8, 9*) and Cel48S (*10*) have been solved with various inhibitors and oligosaccharides, revealing that their active site is located at a junction between a tunnel and a cleft. The tunnel contains binding sites for the incoming cellulose chain and the cleft is the product binding site.

Family 48 cellulases employ concerted one-step general acid/base catalysis to hydrolyze the β-1,4 glycosidic bonds in cellulose (*13*). In this mechanism the catalytic acid serves as the proton donor for the oxygen (O4) in the glycosidic bond undergoing hydrolysis, while the catalytic base activates the catalytic water molecule, which attacks the anomeric carbon (C1), resulting in the inversion of the original stereochemistry. Crystallographic data have made it

possible to establish the catalytic acid in the Cel48 family. It is a glutamic acid residue (E55 in Cel48F, E87 in Cel48S, and E52 in TfCel48A) located in the active site, which makes favorable hydrogen bonding to the O4 atom of the sugar unit in the +1 subsite (*6, 10*). The catalytic base, on the other hand, had not been experimentally established to date because there is no defined sugar residue in the -1 subsite in any of the obtained structures. Combined structural analyses (*6, 8*) and molecular simulations (*14*) suggested an aspartic acid residue (D230 in Cel48F, D255 in Cel48S, and D224 in TfCel48A) to be the most likely candidate to carry out the catalytic base function in Cel48 enzymes. Here, we present strong experimental evidence that D224 is indeed the catalytic base in *T. fusca* TfCel48A.

Methods and Materials

The catalytic domain of *T. fusca* TfCel48A was previously cloned and expressed in *E. coli* in our group (3). Using the QuickChange II-XL Site-directed Mutagenesis Kit (Agilent Technologies) we prepared four mutant versions of the protein: D224N, D224E, D224A, and D224G. The reasons for the chosen mutations are as follows. Asparagine is the closest structural analog of aspartate that is unable to catalyze hydrolysis. Glutamate is the closest structural and functional analog of aspartate that might retain some catalytic activity. Alanine and glycine are the two smallest amino acids, which would provide more flexibility within the active site, needed for the azide rescue experiments described below. All mutants were expressed and purified as previously described for the native enzyme (3). Activity of the enzymes was measured on insoluble phosphoric acid swollen cellulose (PASC) and on 1,4-β-D-celohexaose (G6), the longest soluble cellooligosaccharide. PASC was prepared by treating crystalline cellulose (Sigmacell® 101, Sigma-Aldrich) with phosphoric acid as described in (18) to reduce its crystallinity for easier digestion. Pure G6 was obtained from Megazyme

International. All reactions were prepared by combining TfCel48A with 1.7 mg/ml PASC or 17 µM G6 in 600 µl of 50 mM sodium acetate buffer, pH 5.5. Enzyme concentration was 1 µM on PASC and 0.17 µM on G6. Reactions were incubated at 50 °C for four (G6) or 18 (PASC) hours and soluble products were analyzed via HPLC using Aminex HPX-87P column (Bio-Rad) and water as a mobile phase. As expected, the main products observed were cellobiose and cellotriose. Trace amounts of cellotetraose were also present in some cases. For PASC percent digestion was calculated from the total measure amount of glucose units released from the bulk substrate. For G6 percent digestion was calculated from the amount of G6 left in the reaction buffer at the end of the reaction. Presented data are averages and standard deviation of triplicate measurements.

Results and Discussion

Figure 4.1 shows the activity of the wild type (WT) and mutant enzymes on PASC and G6. On both soluble and insoluble substrates the substitution of aspartate with glutamate (D224E) resulted in significant partial retention of WT activity. All three other mutants showed only residual activity, which is consistent with the presumed catalytic base role of D224. It is interesting to note that the activity of D224E relative to WT is much lower on G6 than on PASC. For cellulases acting on insoluble substrates it is generally considered that access to substrate, not hydrolysis, is the rate limiting step. This is evidenced by the much higher cellulase activities on soluble cellooligosaccharides than on insoluble cellulose. As a mutation of the catalytic base, D224E should only increase the activation barrier of the hydrolysis step without an effect on the ability of TfCel48A to bind cellulose. Hence, it is likely that the effect of the glutamate mutation is partially masked on PASC by the high activation energy required to bind an individual

cellulose chain in the active site, but is fully revealed on a soluble substrate, for which hydrolysis is the rate limiting step.

To further confirm that TfCel48A D224 is the catalytic base and not just an important residue in the catalytic site, we tested the ability of azide to rescue the activity of the mutated versions of the protein (Figure 4.2) (5). For this test, activity assays were carried out as described above, but with the indicated amount of sodium azide added to the reactions. Of the four mutant enzymes only the D224G protein showed an increase in activity on PASC with added azide. The position of the activated water molecule that attacks the anomeric carbon during hydrolysis is critical. We thus presume that the reason for restored activity in the D224G mutant only is due to the greater flexibility of its active site in comparison to the other mutants, which all have more sterically-hindering functional groups. The ability of azide to partially restore enzyme activity when D224 is mutated is consistent with its presumed role as the catalytic base in TfCel48A.

In summary, using site-directed mutagenesis and azide rescue hydrolysis assays we have shown that the *T. fusca* TfCel48A catalytic base is D224. This residue is equivalent to D230 of *C. cellulolyticum* Cel48F and D255 of *C. thermocellum* Cel48S. To our knowledge, this is the first time that the catalytic base in family 48 glicoside hydrolases has been experimentally established.

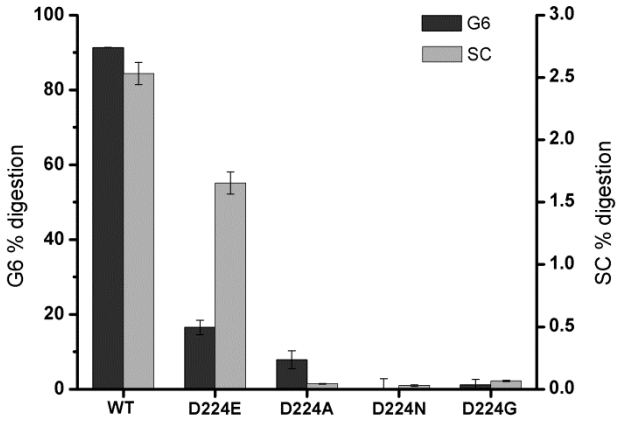


Figure 4.1: Activity of the wild type Cel48A catalytic domain and Cel48A D225 mutants on G6 (soluble) and PASC (insoluble); n = 3.

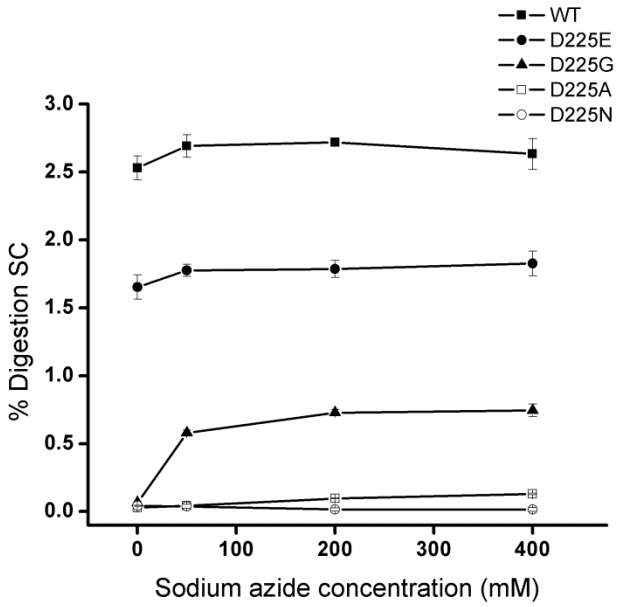


Figure 4.2: Effect of azide addition on the activity of wild type Cel48A catalytic domain and Cel48A D225 mutants on PASC; n = 3

References

1. Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, expression and characterization of a family 48 exocellulase, TfCel48A, from Thermobifida fusca, *Eur J Biochem 267*, 4988-4997.

2. Olson, D. G., Tripathi, S. A., Giannone, R. J., Lo, J., Caiazza, N. C., Hogsett, D. A., Hettich, R. L., Guss, A. M., Dubrovsky, G., and Lynd, L. R. (2010) Deletion of the Cel48S cellulase from Clostridium thermocellum, *Proc Natl Acad Sci U S A 107*, 17727-17732.

3. Berger, E., Zhang, D., Zverlov, V. V., and schwarz, W. H. (2007) Two noncellulosomal cellulases of Clostridium thermocellum, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically, *FEMS Microbiol Lett 268*, 194-201.

4. Zverlov, V. V., Kellermann, J., and schwarz, W. H. (2005) Functional subgenomics of Clostridium thermocellum cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes, *Proteomics 5*, 3646-3653.

5. Wu, J. H. D., Orme-Johnson, W. H., and Demain, A. L. (1988) Two components of an extracellular protein aggregate of Clostridium thermocellum together degrade crystalline cellulose, *Biochemistry 27*, 1703-1709.

6. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 A resolution, *Embo J 17*, 5551-5562.

7. Spiridonov, N. A., and Wilson, D. B. (1998) Regulation of biosynthesis of individual cellulases in Thermomonospora fusca, *J Bacteriol 180*, 3529-3532.

8. Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (2000) Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action, *Biochemistry 39*, 11238-11246.

9. Parsiegla, G., Reverbel, C., Tardif, C., Driguez, H., and Haser, R. (2008) Structures of mutants of cellulase Cel48F of Clostridium cellulolyticum in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action, *J Mol Biol 375*, 499-510.

10. Guimaraes, B. G., Souchon, H., Lytle, B. L., David Wu, J. H., and Alzari, P. M. (2002) The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the Clostridium thermocellum Cellulosome, *J Mol Biol 320*, 587-596.

11. Kruus, K., Wang, W. K., Ching, J., and Wu, J. H. (1995) Exoglucanase activities of the recombinant Clostridium thermocellum CelS, a major cellulosome component, *J Bacteriol 177*, 1641-1644.

12. Reverbel-Leroy, C., Pages, S., Belaich, A., Belaich, J. P., and Tardif, C. (1997) The processive endocellulase CelF, a major component of the Clostridium cellulolyticum cellulosome: purification and characterization of the recombinant form, *J Bacteriol 179*, 46-52.

13. Shen, H., Tomme, P., Meinke, A., Gilkes, N. R., Kilburn, D. G., Warren, R. A., and Miller, R. C., Jr. (1994) Stereochemical course of hydrolysis catalysed by Cellulomonas fimi CenE, a member of a new family of beta-1,4-glucanases, *Biochem Biophys Res Commun 199*, 1223-1228.

14. Saharay, M., Guo, H., and Smith, J. C. (2010) Catalytic mechanism of cellulose degradation by a cellobiohydrolase, CelS, *PLoS One 5*, e12947.

CHAPTER 5

Cel48A from *Thermobifida fusca*: structure and site directed mutagenesis of key residues.

Abstract

We report a new crystal structure of TfCel48A, which shows that this enzyme shares important structural features with the other members of the GH48 family. The active site tunnel entrance of the known GH48 exocellulases is enriched in aromatic residues, which are known to interact well with anhydroglucose units of cellulose. We carried out site-directed mutagenesis studies of these aromatic residues (Y97, F195, Y213, and W313) along with two non-aromatic residues (N212 and S311) also located around the tunnel entrance as well as a W315 residue inside the active site tunnel. Only the aromatic residues located around the tunnel entrance appear to be important for the ability of TfCel48A to access individual cellulose chains on bacterial cellulose (BC), a crystalline substrate. Both Trp residues are also important for processivity of TfCel48A on BC and phosphoric acid swollen cellulose (PASC). On BC reduced processivity correlates with reduced enzyme activity, but the reverse is true on PASC. Presumably, higher density of available cellulose chain ends and the amorphous nature of PASC explain the increased initial activity of mutants with lower processivity. Our data also suggest, however, that processivity is more important for the overall ability of the enzymes to digest PASC as judged by the curvature of the time course activity assay profiles.

Introduction

The rate of crystalline cellulose hydrolysis is much lower than that of soluble cellooligosaccharides and it is widely believed that access to substrate – the individual cellulose chains – is the rate limiting step in enzymatic digestion of cellulose. Recent experimental evidence, including data presented in Chapter 3, further supports this hypothesis (*1, 2*). It is likely, therefore, that at least some cellulases have evolved residues whose primary function is to enhance the ability of the enzyme to separate individual cellulose chains away from the bulk substrate and thread them to the active site.

Several structures of three unique GH48 enzymes – Cel48F from *Clostridium cellulolyticum(3),* Cel48S from *C. thermocellum(4)* and GH48A from *Hahella chejuensis(5)* – have been published to date. All three enzymes share the same (alpha/alpha)$_6$ barrel fold and an almost identical active site tunnel that can accommodate seven glucose units at subsites -7 to -1, followed by an open product binding site, best adapted for cellobiose. Here we report the crystal structure of another GH48 exocellulase, Cel48A from *Thermobifida fusca*, a model cellulolytic bacterium. The structure was obtained with X-ray diffraction using the TfCel48A catalytic base mutant D224N (*6*) complexed with cellobiose and cellohexaose. In order to better understand the mechanisms by which cellulases are able to hydrolyze crystalline substrates we generated and tested point mutants of several aromatic and two conserved non-aromatic residues near the entrance of the active site tunnel. Our results suggest that aromatic residues near the binding site of cellulases play an important role in the ability of these enzymes to access individual cellulose chains when acting on microcrystalline cellulose.

TfCel48A crystallization, structure determination, structure analysis were done at the National Renewable Energy Laboratory by Markus Alahuhta, Roman Brunecky, and Vladimir Lunin. Modeling of the tunnel entrance was done by Mo Chen in John Brady's laboratory at Cornell University.

Methods and Materials

*Substrates***.**

Bacterial cellulose (BC) was a gift from Monsanto. BC cake was washed three times with deionized (DI) water by centrifugation and resuspended in DI water with 0.04% sodium azide (Sigma-Aldrich). Concentration was determined as dry weight per volume. Phosphoric acid swollen cellulose (PASC) was prepared from Avicel® powder (PH-105; FMC Corporation,

Philadelphia, PA, USA) using procedures described in (*7*) and was stored in DI water with 0.04% sodium azide.

*Enzymes***.**

The CD of TfCel48A-D224N for structure analysis was expressed and purified as previously described in (*8*). WT and all other mutant versions of TfCel48A were expressed and purified as previously described in Chapter 2. TfCel48A mutants were generated using Agilent Quickchange® II XL Site Directed Mutagenesis Kit following the manufacturer's instructions. All mutations and sequences were verified by Sanger sequencing. All enzyme concentrations were determined by spectroscopy using NanoDrop® 1000 spectrophotometer.

*Crystallization***.**

TfCel48A-D224N crystals were obtained with sitting drop vapor diffusion using a 96-well plate with Crystal Screen HT from Hampton Research (Aliso Viejo, CA). 50 µL of well solution was added to the reservoirs and drops were made with 0.2 µL of well solution and 0.2 µL of protein solution using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, CA). The crystals were grown at 20°C with 0.1 M sodium cacodylate trihydrate pH 6.5, 18% (w/v) polyethylene glycol 8000 and 0.2 M zinc acetate dihydrate as the well solution. The protein solution contained 1.2 mg/mL of protein, 20 mM acetate pH 5.0, 100 mM NaCL, 5 mM $CaCl_2$, 5 mM cellohexaose and 1 mM cellobiose.

*Data collection and processing***.**

The TfCel48A-D224N crystal was flash frozen in a nitrogen gas stream at 100 K before data collection. The crystallization solution with 20% (v/v) ethylene glycol and glycerol was used for freezing the crystal. Data was collected in an in-house Bruker X8 MicroStar X-Ray

generator with Helios mirrors and Bruker Platinum 135 CCD detector. Data were indexed and processed with the Bruker Suite of programs version 2011.2-0 (Bruker AXS, Madison, WI).

Structure solution and refinement.

Intensities were converted into structure factors and 5% of the reflections were flagged for Rfree calculations using programs SCALEPACK2MTZ, ctruncate, MTZDUMP, Unique, CAD, FREERFLAG and MTZUTILS from the CCP4 package of programs (*9*). The program MOLREP version 11.0.05 (*5*)) as the molecular replacement. The best solution was achieved using REFMAC5 (*5*) version 5.7.0032 and Coot (*10*) version 5.7.0032 and Coot (*11*) version 0.6.2. The MOLPROBITY method (*12*) was used to analyze the Ramachandran plot and root mean square deviations (rmsd) of bond lengths and angles were calculated from ideal values of Engh and Huber stereochemical parameters (*13*). Wilson B-factor was calculated using ctruncate version 1.5.1. Average B-factors, were calculated using program ICM version 3.7-2a (Molsoft LLC, La Jolla, CA). The data collection and refinement statistics are shown in Table 5.1.

Structure Analysis

Programs Coot, PyMOL (http://www.pymol.org) and ICM (http://www.molsoft.com) were used for comparing and analyzing structures. This structure has been deposited to the protein data bank (PDB; www.rcsb.org) with entry code 4JJJ.

Time course and endpoint activity assays.

All reactions were conducted in triplicate in Eppendorf 2 mL Protein LoBind plastic tubes. 1.5 mg substrate was combined with 0.1 µM enzyme in 0.6 mL 50 mM sodium acetate buffer, pH 5.5. Only buffer and substrate were combined for negative controls. Upon mixing, BC and PASC reactions were immediately placed in a 50 °C (unless otherwise indicated) water bath.

Samples in triplicate were removed at the given time points and placed on dry ice to stop the reaction. Frozen samples were later placed in a boiling bath for 10 minutes in order to denature the enzyme. It was verified experimentally that boiling does not alter soluble sugar profiles detected by High Performance Liquid Chromatography (HPLC). The remaining substrate was removed using Corning® Spin-X® Centrifuge tube filters and the soluble sugar concentrations were measured using a Shimadzu HPLC system fitted with a Bio-Rad Aminex® HPX-87P analytical column and a refractive index detector. The mobile phase was Milli-Q water at a flow rate of 0.6 ml/min. Sample injection (25-50 µl volume) was performed by an autosampler installed on the instrument.

Apparent K_M assays.

0.1 µM TfCel48A was combined with the indicated amount of PASC in 0.6 ml of 50 mM sodium acetate, pH 5.5 and reactions were incubated in a 50 °C water bath for one hour. Samples were boiled for five minutes to denature the enzyme and were processed and analyzed by HPLC in the same manner as described for the time course assays.

Data analysis.

HPLC data were processed with OriginPro 8. Product identities and concentrations were determined by Gaussian peak fitting, using standard solutions with known concentrations of soluble cellooligosaccharides for reference. Soluble sugar concentrations at time zero were subtracted from all of the subsequently obtained concentrations. Soluble sugar produced upon initial mixing of the enzyme and the substrate is primarily due to the burst activity, as described in (*14, 15*), whereas the model used here is concerned with the digestion of the more recalcitrant portions of cellulose. Apparent K_M and the *A* and *b* parameter values of time course profiles were

determined using the nonlinear least squares fit of the Michaelis-Menten expression ($V_{max}[S]/(K_M + [S])$) and Eqn. 5.1 respectively.

Results and Discussion

The crystal structure of T. fusca Cel48A

The structure of TfCel48A-D224N was refined to a resolution of 1.6 Å with R and Rfree of 0.121 and 0.164, respectively. There is one molecule in the asymmetric unit in complex with a cellobiose and a cellohexaose molecule (Figure 5.1). It has an $(alpha/alpha)_6$ barrel fold with several calcium, sodium, and zinc atoms and one ethylene glycol and one acetate molecule bound on the surface. The mutated 224 residue is well defined in the electron density.

Structural comparison

Pair-wise secondary-structure matching of structures with at least 70% secondary structure similarity by PDBfold (*16*) found 22 unique structural matches for GH48 from the protein data bank. Out of these only three were unique. Most were CcCel48F and CtCel48S structures with one GH48 from *H. chejuensis*. TfCel48A has 54% sequence identity with CtCel48S (*4*) and 56% sequence identity with CcCel48F (*3*) and *H. chejuensis* GH48 . The Cα root mean square deviations of all similar structures varied between 0.85 Å and 1.01 Å showing that the overall backbone of all of the known GH48 structures is similar.

Closer inspection of the active site tunnels of the other three unique GH48 structures (PDB entry codes 1FCE, 1L2A and 4FUS) with TfCel48A shows it to be highly conserved. The catalytic glutamate and the cellobiose and cellohexaose molecules superimpose almost perfectly with those of the other structures. When residues located within five angstroms from cellobiose and cellohexaose are compared, only two differences stand out. Ile106 is alanine and a

tryptophan at position 406 is tyrosine in the other GH48 structures. Both changes are functionally similar further emphasizing the highly conserved nature of the known GH48s.

The asparagine residue of the D224N mutation adopts a slightly different conformation compared to the aspartate residues of the other GH48 structures. The mutation makes it impossible for this residue to function as the catalytic base and causes it to lose its hydrogen bond with the main chain oxygen of A222. N224 instead forms a hydrogen bond with R424, consequently changing its orientation.

Cel48A activity and processivity on crystalline and amorphous cellulose.

The kinetic model of enzymatic cellulose hydrolysis described in Chapter 2 was used in this study and a brief review of the model follows. The model relies on two parameters to fit experimentally obtained time course profiles of individual cellulases and their mixtures acting on insoluble cellulose. The equation used in the model is based on a pseudo zero order Michaelis-Menten scheme (*17*), but contains a time (digestion) dependent activity coefficient rather than a constant to account for the continuous decrease in activity typically observed for cellulose hydrolysis. It can be expressed either as product formation over time, or as % digestion over time. The latter is used in this manuscript:

$$X = At^b \qquad 0 \leq b \leq 1 \tag{5.1}$$

where X is % digestion of the substrate, A is the net activity of the added enzyme, and b is an intrinsic constant, which quantifies the curvature of the time course profile and reflects the ability of a given enzyme to degrade a given substrate. Parameter values are obtained by fitting experimental time course data to Eqn. 1 and the model is applicable under conditions of low to medium-low enzyme loads. The A parameter is a product of specific activity and the concentration of the productively bound enzyme. It is strongly dependent on the total added

enzyme concentration and is exponentially dependent on temperature, following the Arrhenius equation. As an intrinsic constant, *b* parameter of individual cellulases is not dependent on enzyme concentration and shows only a slight dependence on temperature. The theoretical limits of *b* are 0, in which case no products would be generated over time and 1, in which case the rate coefficient is constant over time, as is the case in classical kinetics. Consequently, cellulases with higher *b* values on a given substrate are more effective at overcoming the recalcitrance of that substrate.

By itself, TfCel48A is the least effective of *T. fusca* cellulases in the digestion of insoluble cellulose (*8*). This is reflected by its low *A* and *b* parameter values obtained from the time course profiles on BC (bacterial cellulose) and PASC (phosphoric acid swollen cellulose), shown in Figure 5.2a. On BC, which has high crystallinity (*18*), its *A* and *b* values are 0.45 and 0.34 respectively. PASC is a mostly amorphous form of cellulose, obtained by treatment of crystalline cellulose with phosphoric acid (*7, 19*). It is more readily degraded than crystalline cellulose and this is reflected in the higher TfCel48A *A* and *b* values of 1.51 and 0.41 respectively. The higher *b* value on PASC indicates that this substrate is more accessible for TfCel48A, which results in a slower decline of the digestion rate over time when compared to BC hydrolysis. The higher *A* value observed on PASC may be due to an increase in either or both, the productively bound fraction of the total added enzyme and the specific activity of TfCel48A acting on less tightly packed cellulose chains.

Processivity is an important consideration in understanding the mechanisms of cellulose hydrolysis. If access to substrate is indeed the rate limiting step on crystalline cellulose, it should be advantageous for a cellulase to carry out multiple cleavage steps on the same chain that has been threaded in the active site tunnel. This is not necessarily the case on soluble and amorphous

substrates (*20*). It is not practical to measure processivity directly on unmodified insoluble cellulose such as BC and PASC due to technological limitations, and for this reason relative processivity values have been employed in literature. For exocellulases processivity is typically determined by measuring the relative concentrations of glucose (G1), cellobiose (G2), and cellotriose (G3) (*21-23*). Depending on the mode of initial binding of the cellulose chain in the active site, the first cleavage product may be G1, G2, or G3, but only G2, the repeating unit of a cellulose chain, is produced for all subsequent processive cleavages. In the case of TfCel48A acting on BC or PASC, the initial cleavage appears to be only either G3 or G2, as no G1 is ever observed. In addition, at low enzyme concentration, a trace amount of cellotetraose (G4) is also produced, which is most likely the result of a skipping event during procession of the enzyme. G4 is relatively slowly cleaved by the enzyme to produce two G2 units. As a result, relative processivity (*P*) of TfCel48A is given here as:

$$P = \frac{[G4]+[G2]}{[G3]} \tag{5.2}$$

where brackets indicate product concentration.

The effect of temperature on the activity of TfCel48A was previously studied by our group, with optimal temperature determined to be around 50 °C (*8*). The effect on processivity, however, was not determined at that time. As Figure 5.2b illustrates, when the temperature of BC digestion by TfCel48A is increased from 10 to 50 °C, the ratio of the sum of G4 and G2 to G3 is decreased from ca. 18 to 14, indicating a decrease in processivity. This is most likely due to an increase with temperature in the value of k_{off}, which governs the binding affinity of TfCel48A to the cellulose chain inside the active site tunnel.

The role of aromatic residues in binding chain ends.

Aromatic residues play an important role in the ability of carbohydrate-active enzymes to bind via stacking interactions with pyranose rings (*24*) and their importance for cellulose binding by cellulases has been demonstrated in literature (*20, 21, 25, 26*). The tunnel entrance of TfCel48A and other GH48 cellulases is enriched in aromatic residues and it is possible that these residues are important for the ability of these enzymes to separate and thread individual cellulose chains from the bulk substrates. In order to test this hypothesis, we prepared alanine point mutations of the four aromatic residues in TfCel48A, as shown in Figure 5.3a: Y97, F195, Y213, and W313. Another aromatic residue located inside the tunnel, W315 was mutated in order to compare its role with that of the residues immediately outside the tunnel. As shown in Figure 5.3b, both Trp residues are in line with each other and form stacking interactions with the cellulose chain located in the active site tunnel. Two non-aromatic residues, N212 and S311, which are located next to Y211 and W313 respectively, were also mutated. Finally, double mutations – Y213A-S311A and Y213A-W313A – were generated to see whether the effect of the tested mutations on TfCel48A properties is additive. All mutants were analyzed by native polyacrylamide gel electrophoresis to verify that the mutations did not compromise protein folding or stability (Figure 5.4).

Soluble cellooligosaccharides such as cellopentaose (G5) can serve as a proxy for loose chain ends on bulk cellulose. Activity of TfCel48A mutants on G5 is shown in Figure 5.5. Reduced activity is observed for Y97A, F195A, and Y213A single mutants as well as for both double mutants, which contain the Y213A mutation. This result suggests that these residues may be important for the initial binding of TfCel48A to soluble chain ends.

The role of aromatic residues in the hydrolysis of BC.

Time course profiles of BC hydrolysis were fit with Equation 1 and the obtained parameter values are plotted in Figure 5.6. A relatively small, but consistent reduction of the *b* value (0.26 – 0.29) is apparent for all point mutants of the aromatic residues located around the active site tunnel entrance. WT value (0.34) is retained for both non-aromatic residue mutations N212A (0.36) and S311A (0.34). The effect of the aromatic residue mutations is more pronounced in the double mutant Y213A-W313A, as evidenced by its significantly lower *b* value (0.14). On the other hand, the value of the double mutant Y213A-S311A (0.30) is similar to that observed for the point mutant Y213A (0.29). Of particular importance is the observation that the mutation of W315 which is aromatic, but is located inside the tunnel, does not seem to affect the *b* value (0.33) of the protein, as we previously showed for slightly different enzyme and substrate concentrations (see Chapter 2). Together, these results suggest that of the tested residues, the aromatic amino acids located around the active site tunnel entrance are particularly important for the ability of TfCel48A to access the individual cellulose chains packed in BC. This is consistent with the demonstrated ability of aromatic residues to interact with glucose units of cellulose (*24*) and with the location of these residues near the active site tunnel (Figure 5.3a).

The *A* parameter in Equation 5.1 is the net activity of the added enzyme, and changes in its value may be indicative of the change in the specific activity of the enzyme or the productively bound enzyme fraction or both. Compared to WT (0.45), the *A* value of the mutants, (Figure 5.6b) is slightly reduced for Y97A (0.42), F195A (0.41), and Y213A (0.41), all of which are located near the tunnel entrance, but according to structural data do not interact with the cellulose chain inside the tunnel (Figure 5.3). The slight reduction in the *A* value of these mutants may be indicative of a smaller than WT fraction of productively bound enzyme, which

would be consistent with their reduced *b* parameter values. The *A* value of W313A and W315A is significantly lower than that of WT, and it correlates with a lower processivity of these mutants, as quantified with Equation 5.2 (Figure 5.6c). Lower processivity for these mutants is expected from the structural data, which indicates that both Trp residues form stacking interactions with the cellulose chain inside the active site tunnel (Figure 5.3b). The lower *A* values may be the result of both a smaller than WT fraction of productively bound enzyme and a lower specific activity due to a slower procession along the chain as a result of the introduced mutations. The fact that none of the other mutations result in lower processivity of TfCel48A suggests that the other tested residues do not interact with the cellulose chain once it is inside the active site tunnel.

Effect of the mutated residues on the digestion of PASC.

PASC is a mostly amorphous form of cellulose with different properties compared to highly crystalline substrates such as BC. It is more digestible by cellulases due to its smaller particle size, lack of supramolecular structure, and lower degree of polymerization (*7*), and WT TfCel48A is about five times more active on PASC than on BC (Figure 5.2). In order to further understand the function of the mutated residues, similar time course experiments were carried out on PASC to determine the effect of the mutations on the parameter values of Equation 5.2, shown in Figure 5.7. Of the tested mutants, only W313A and W315A have a significant effect on the value of the *b* parameter (0.34 and 0.33 respectively) in comparison to WT (0.41). This is interesting because W315A did not affect the *b* value on BC and it suggests that the basis of rate declines is different on BC and PASC. As a regenerated form of cellulose with no supramolecular structure (*7*), it is likely that PASC contains many obstacles due to crisscrossing of individual cellulose chains and a lack of defined fibers. The reduced *b* value of both Trp

mutants correlates with reduced processivity. Published modeling and experimental data show that the removal of aromatic residues that interact with the cellulose chain in the tunnel of exocellulases can reduce the binding affinity of the enzyme (*20, 25, 26*). This appears to be the case with the W313A and W315A mutations, as further supported by their higher apparent K_M values on PASC (1.7 and 2.4 mg/ml PASC respectively, compared to 0.7 for WT; Figure 5.8). Inherent processivity of exocellulases is governed by the k_{off} constant, which in turn is related to their binding affinity for the cellulose chain undergoing hydrolysis. Thus, affinity for the substrate inside the active site tunnel may play a dominant role in the ability of an exocellulase to effectively digest PASC and for this reason the *b* value of TfCel48A correlates with processivity. In turn, this suggests that unlike BC, digestibility of PASC by an exocellulase is related to the ability of the enzyme to overcome surface obstacles before dissociating from the chain.

The *A* value of the aromatic residue mutants (1.58-2.19) is slightly to moderately higher than that of WT (1.51). This is somewhat surprising for the Y97A, F195A, Y213A, and Y213A-S311A mutants, which have WT processivity and slightly decreased *b* values. One possible explanation for this is that the removal of these residues leads to a small increase in TfCel48A specific activity due to a weaker interaction with the cellulose chain end upon initial binding without compromising the ability of the enzyme to bind productively to the substrate. In the case of both the W313A and W315A mutants, the increased *A* value (2.06 and 2.19) correlates with, and can be explained by their lower processivity. Since lower processivity indicates weaker binding of the enzyme to the cellulose chain undergoing hydrolysis it is possible that these mutants are able to process faster along that chain until they dissociate. Given the lower degree of polymerization and higher accessibility of PASC (*7*) higher frequency of dissociation from the chains is not necessarily detrimental to the overall rate of digestion, as a dissociated enzyme

should be able to rapidly find another productive binding site to continue hydrolysis. This is consistent with other published data on the effect of aromatic residue mutations in processive cellulases and chitinases acting on amorphous substrates (*20*).

Conclusions.

We have determined the X-ray crystal structure of the *T. fusca* TfCel48A D224N mutant in complex with cellobiose and cellohexaose, and structural comparisons show high similarity with the active site tunnels of other known GH48 enzymes. Five aromatic residues around the TfCel48A active site tunnel entrance and one inside the tunnel were mutated and the mutated enzymes were tested on cellopentaose, BC, and PASC. According to data obtained on cellopentaose Y97, F195, and Y213, all of which are located around the tunnel entrance, but are not bound to the cellulose chain undergoing hydrolysis, may play a role in helping thread cellulose chain ends into the active site tunnel. Both Trp residues – W313 and W315 – which are located at the mouth and inside the active site tunnel respectively and interact with the cellulose chain undergoing hydrolysis, are important for processivity of TfCel48A on BC, PASC, and probably other substrates. All aromatic residues located around the tunnel entrance seem to be important for the ability of TfCel48A to digest BC presumably due to the affinity of aromatic residues for the glucose units of cellulose. On the other hand, the ability of TfCel48A to digest PASC appears to be more dependent on its processivity, which is governed by the strength of its interaction with the cellulose chain inside its active site tunnel. The lack of conservation of most of the residues around the active site tunnel including three of the aromatic residues that are required for crystalline cellulose activity is in stark contrast to the conservation of the active site residues in the tunnel. One possible explanation for this is that the tunnel entrance residues have been selected for the types of cell walls utilized by *T. fusca* so that each

organism has evolved a set of such residues that is optimal for its substrates. If this is true then it should be possible to produce engineered variants that will have higher activity on specific pretreated biomass substrates.

Table 5.1: X-ray data collection and refinement statistics. Statistics for the highest resolution bin are in parenthesis.

Data collection	
Space group	P21
Unit cell, Å, °	a= 56.458, b= 88.494, c = 66.753 β = 113.7
Wavelength, Å	1.54178
Temperature (K)	100
Resolution, Å	25-1.6 (1.7-1.6)
Unique reflections	78727 (12722)
R_{int} †	0.089 (0.251)
Average redundancy	7.93 (3.32)
$\langle I \rangle / \langle \sigma(I) \rangle$	15.10 (3.81)
Completeness, %	99.4 (96.9)
Refinement	
Resolution, Å	25-1.6 (1.64-1.60)
R/R_{free}	0.121 (0.190)/ 0.164 (0.244)
Protein atoms	5415
Water molecules	1175
Other atoms	160
RMSD from ideal bond length, Å#	0.023
RMSD from ideal bond angles, ° #	2.152
Wilson B-factor	7.25
Average B-factor for protein atoms, $Å^2$	10.2
Average B-factor for water molecules, $Å^2$	23.6
Ramachandran plot statistics, %*	
Allowed	99.9
Favored	97.5
Outliers	1 (Ala41)

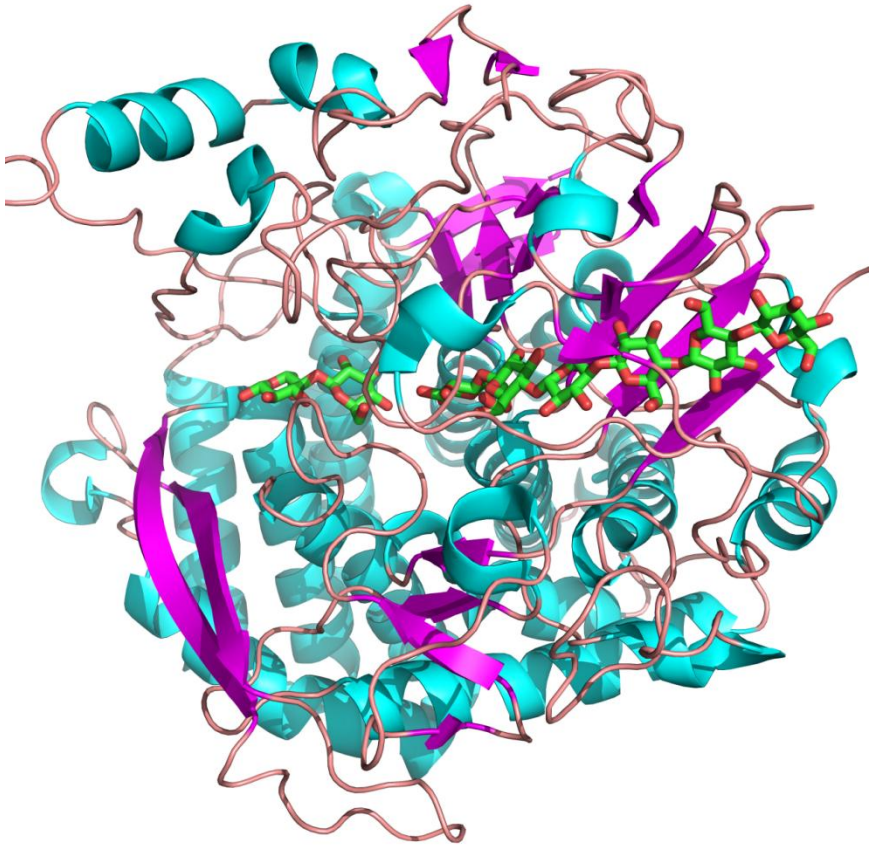


Figure 5.1: Overall structure of T. fusca Cel48A. Cellobiose (left) and cellohexaose (right) are shown as sticks with green carbons and red oxygen atoms

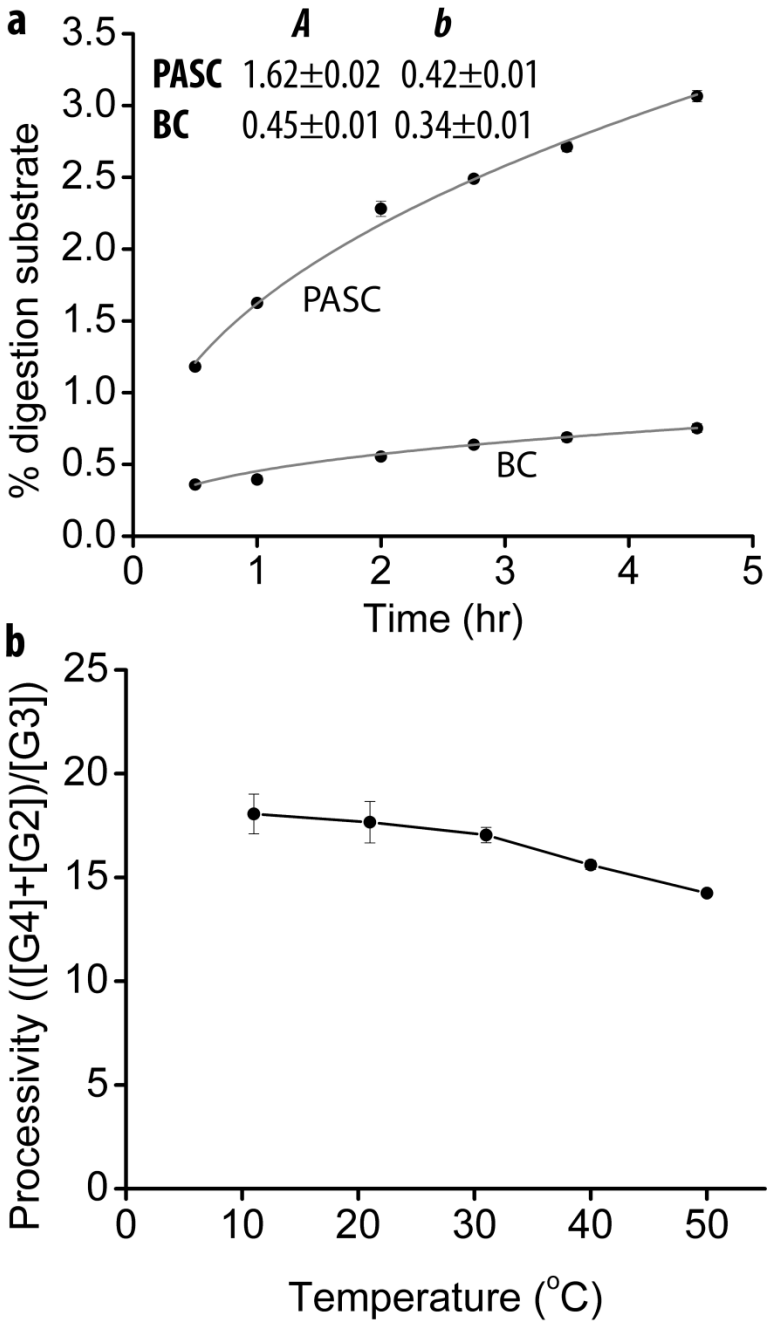


Figure 5.2: Activity and processivity of wild type TfCel48A. a) Time course profiles of 0.1 µM TfCel48A on 1.5 mg BC and PASC with corresponding A and b parameter values from Equation 1. b) Relative processivity (Equation 5.2) of TfCel48A on BC as a function of temperature.

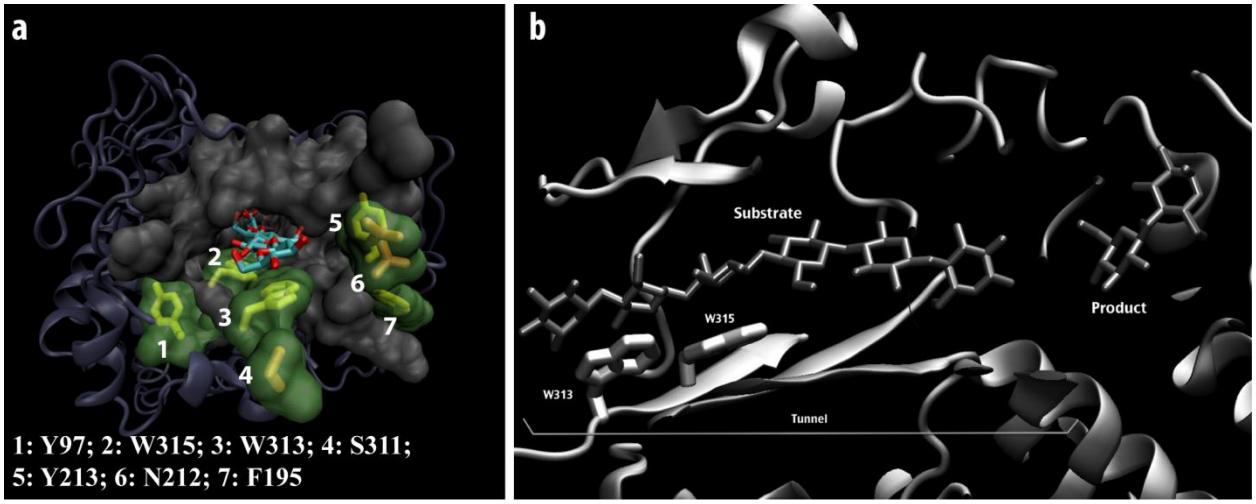


Figure 5.3: TfCel48A active site tunnel with substrate. Highlighted residues were mutated individually or in pairs to alanines. a) Residues around the tunnel entrance are fully or partially exposed on the surface of the enzyme. b) W313 and W315 are located at the mouth and inside of the tunnel respectively and both interact with the cellulose chain bound in the tunnel.

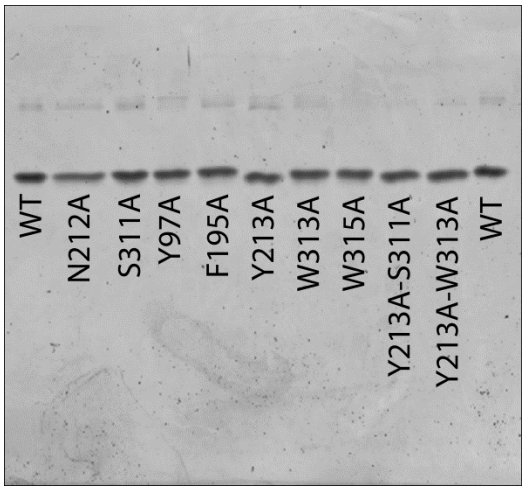


Figure 5.4: Native polyacrylamide gel of WT and mutant versions of TfCel48A. All mutants run the same as WT, indicating that the mutations did not cause apparent folding or stability issues.

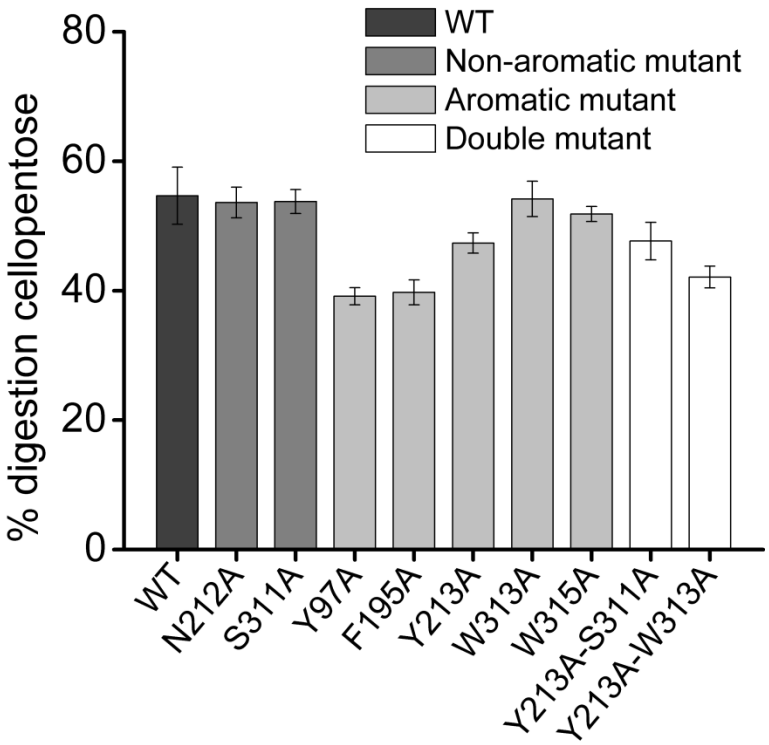


Figure 5.5: Activity of TfCel48A wild type and mutants on cellopentaose (G5), measured as % digestion. All points were obtained in triplicate and error bars represent one standard deviation.

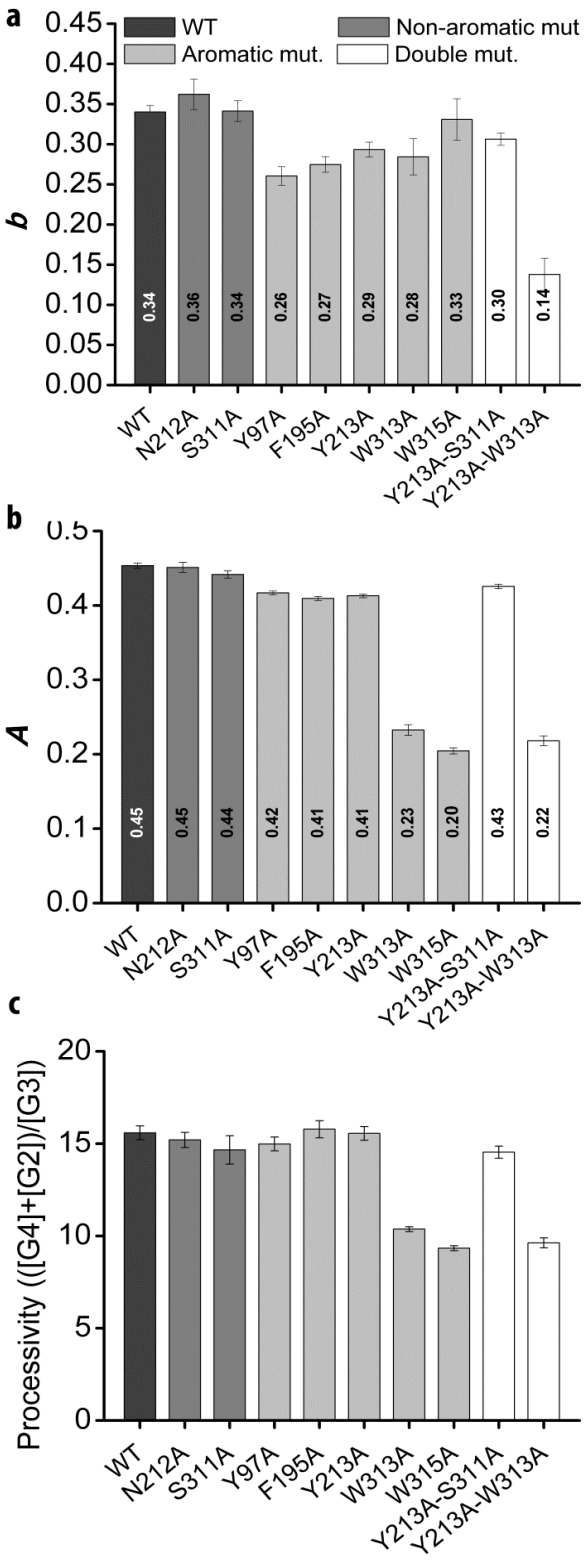


Figure 5.5: Activity of 0.1 µM TfCel48A wild type and mutants on 1.5 mg BC. b, the hydrolysis power factor (a) and A, the net activity (b) were obtained by fitting the time course of each protein variant with Equation 1. All data points were obtained in triplicate and error bars represent standard error of the fit.

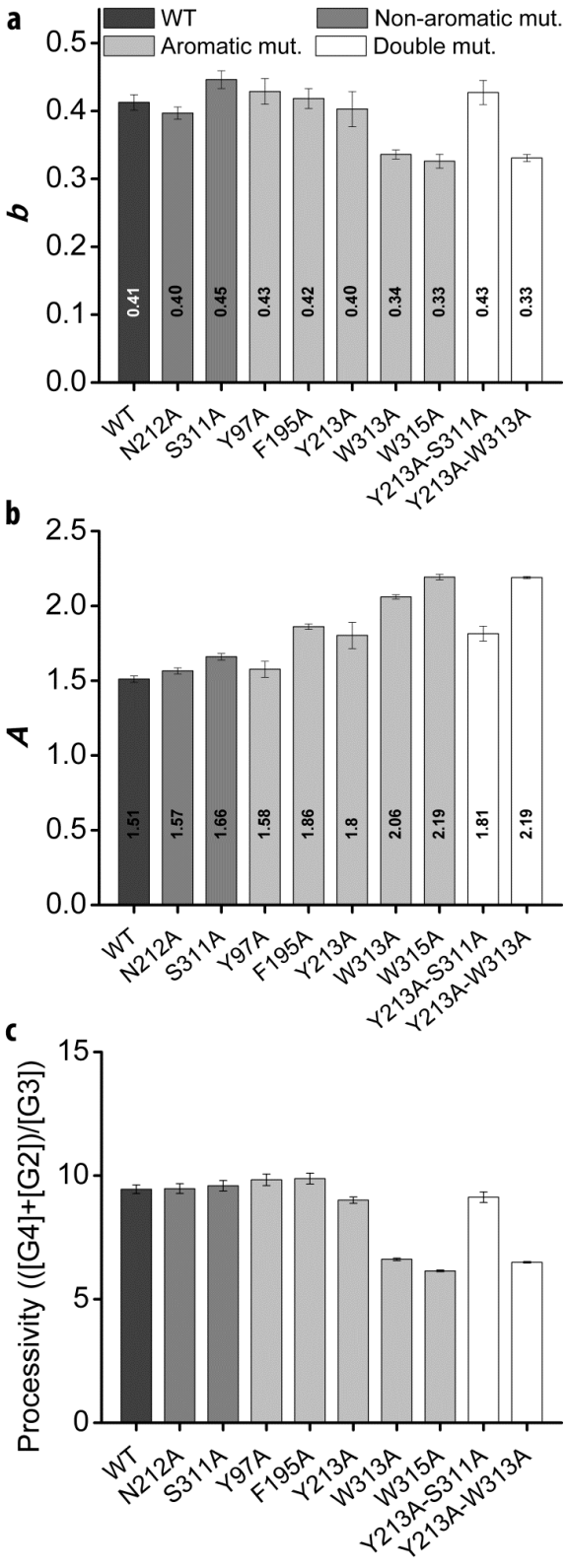


Figure 5.6: Activity of 0.1 µM TfCel48A wild type and mutants on 1.5 mg PASC. *b*, the hydrolysis power factor (a) and *A*, the net activity (b) were obtained by fitting the time course of each protein variant on PASC with Equation 1. All data points were obtained in triplicate and error bars represent standard error of the fit.

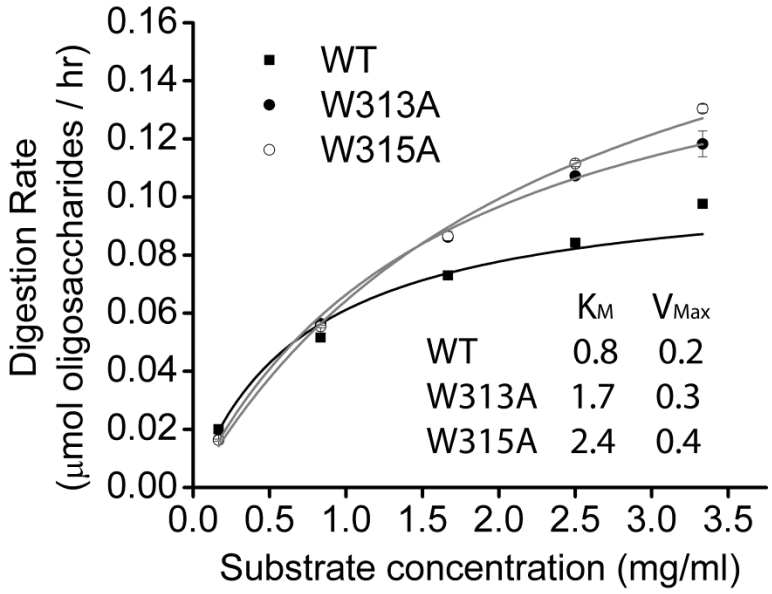


Figure 5.7: Apparent Km of TfCel48A WT and the mutants W313A and W315A. Both Trp mutations result in a higher apparent Km and Vmax values compared to WT.

References

1. Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2011) Determination of the molecular states of the processive endocellulase Thermobifida fusca Cel9A during crystalline cellulose depolymerization, *Biotechnol Bioeng*.

2. Fox, J. M., Levine, S. E., Clark, D. S., and Blanch, H. W. (2012) Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes, *Biochemistry 51*, 442-452.

3. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 A resolution, *EMBO J 17*, 5551-5562.

4. Guimaraes, B. G., Souchon, H., Lytle, B. L., David Wu, J. H., and Alzari, P. M. (2002) The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major

enzymatic component of the Clostridium thermocellum Cellulosome, *J Mol Biol 320*, 587-596.

5. Sukharnikov, L. O., Alahuhta, M., Brunecky, R., Upadhyay, A., Himmel, M. E., Lunin, V. V., and Zhulin, I. B. (2012) Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48, *J Biol Chem 287*, 41068-41077.
6. Kostylev, M., and Wilson, D. B. (2011) Determination of the catalytic base in family 48 glycosyl hydrolases, *Appl Environ Microbiol 77*, 6274-6276.
7. Zhang, Y. H., Cui, J., Lynd, L. R., and Kuang, L. R. (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure, *Biomacromolecules 7*, 644-648.
8. Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca, *Eur J Biochem 267*, 4988-4997.
9. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr D Biol Crystallogr 67*, 235-242.
10. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr D Biol Crystallogr 67*, 355-367.
11. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

12. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12-21.

13. Engh, R. A., and Huber, R. (1991) Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement, *Acta Crystallographica Section A 47*, 392-400.

14. Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K. C., Borch, K., and Westh, P. (2011) A kinetic model for the burst phase of processive cellulases, *Febs J 278*, 1547-1560.

15. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., and Westh, P. (2012) Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A, *J Biol Chem 287*, 18451-18458.

16. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallographica Section D-Biological Crystallography 60*, 2256-2268.

17. Michaelis, L., and Menten, M. L. (1913) Die kinetik der invertinwirkung., *Biochem Z. 49*, 333-369.

18. Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A., and Johnson, D. K. (2010) Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance, *Biotechnol Biofuels 3*, 10.

19. Walseth, C. S. (1952) The Influence of the Fine Structure of Cellulose on the Action of Cellulases, *Tappi 35*, 233-238.

20. Horn, S. J., Sikorski, P., Cederkvist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M., and Eijsink, V. G. (2006) Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides, *Proc Natl Acad Sci U S A 103*, 18089-18094.

21. Vuong, T. V., and Wilson, D. B. (2009) Processivity, synergism, and substrate specificity of Thermobifida fusca Cel6B, *Appl Environ Microbiol 75*, 6655-6661.

22. Medve, J., Karlsson, J., Lee, D., and Tjerneld, F. (1998) Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: adsorption, sugar production pattern, and synergism of the enzymes, *Biotechnol Bioeng 59*, 621-634.

23. von Ossowski, I., Stahlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., Zhao, Y. X., Driguez, H., Claeyssens, M., Sinnott, M. L., and Teeri, T. T. (2003) Engineering the exo-loop of Trichoderma reesei cellobiohydrolase, Ce17A. A comparison with Phanerochaete chrysosporium Cel7D, *J Mol Biol 333*, 817-829.

24. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis, *J Biol Chem 266*, 5202-5219.

25. Poole, D. M., Hazlewood, G. P., Huskisson, N. S., Virden, R., and Gilbert, H. J. (1993) The role of conserved tryptophan residues in the interaction of a bacterial cellulose binding domain with its ligand, *FEMS Microbiol Lett 106*, 77-83.

26. Payne, C. M., Bomble, Y. J., Taylor, C. B., McCabe, C., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011) Multiple functions of aromatic-carbohydrate interactions in a

processive cellulase examined with molecular simulation, *J Biol Chem 286*, 41028-41035.

CHAPTER 6

A Distinct Model of Synergism between a Processive Endocellulase TfCel9A and an Exocellulase TfCel48A from *Thermobifida fusca.*

Abstract

TfCel9A and TfCel48A are two cellulases with different properties and have previously been shown to synergize well with each other. TfCel9A is a processive endocellulase with relatively high activity on crystalline cellulose. TfCel48A is a reducing end directed exocellulase with very low activity on crystalline cellulose. Neither enzyme fits its respective role in the classical model of endo-exo synergism of enzymatic cellulose digestion. Using time course, endpoint, and sequential addition activity assays we propose a model of synergistic cooperation between the two cellulases. TfCel9A appears to be most effective on fresh bacterial cellulose with a presumably uniform surface at the molecular level. Its processive activity likely erodes the surface and thus reduces its own activity. TfCel48A is able to hydrolyze efficiently the TfCel9A-modified substrate and replenish the uniform surface required by TfCel9A, creating a feedback mechanism. The model of synergistic interactions is comparable to an earlier proposed model for *Trichoderma reesei* Cel7A and Cel7B, but the roles of endo- and exo-cellulases are reversed, which suggests that bacteria and fungi may have evolved different approaches to efficient biomass degradation.

Introduction

Studies of synergistic interactions in enzymatic cellulose hydrolysis can contribute to the development of more complete models of synergism than are presently available. Here we report data for synergistic interactions between TfCel9A and TfCel48A, two cellulases secreted by a model cellulolytic bacterium *Thermobifida fusca*. TfCel9A is a processive endocellulase, which contains two cellulose binding modules, CBM2 and CBM3c (*1*) and is relatively effective as a single cellulase in the digestion of crystalline cellulose (*2*). CBM2 is attached via a flexible linker (which includes a Fibronectin III domain) to CBM3c. CBM3c is in turn connected almost rigidly to the GH9 catalytic domain (*3*) and is believed to be responsible for the processivity and

ability of TfCel9A to digest effectively crystalline cellulose (*1, 3*). TfCel48A is a reducing end directed exocellulase with very poor activity on crystalline cellulose (*4*). Despite its low activity TfCel48A makes up about a third of total secreted cellulases when *T. fusca* is grown on cellulose (*5*) and it has been shown to interact synergistically with other *T. fusca* cellulases including TfCel9A (*2, 4*). Given that TfCel9A is processive and is the most effective *T. fusca* cellulase on crystalline substrate (*2*), it is unlikely to serve the typical role assigned to endocellulases in the endo-exo model of synergism. In addition, very low activity of TfCel48A on crystalline cellulose also makes it incompatible with the presumed role of exocellulases in the endo-exo model. Hence, it is likely that TfCel48A and TfCel9A employ a distinct mechanism of synergism and here we propose a model for their cooperation.

Materials and Methods

*Substrates and enzymes***.**

Bacterial cellulose (BC) was a gift from Monsanto. BC cake was washed three times with deionized (DI) water by centrifugation and resuspended in DI water with 0.04% sodium azide (Sigma-Aldrich). Concentration was determined as dry weight per volume. Recombinant TfCel9A and TfCel48A were expressed in *Escherichia coli* BL21 cells and purified as previously described in (*1*) and Chapter 2 respectively.

Time course activity assays.

All reactions were conducted in triplicate in Eppendorf 2 mL Protein LoBind plastic tubes. 1 mg substrate was combined with 83 nM TfCel9A and/or 750 nM TfCel48A in 0.6 mL 50 mM sodium acetate buffer, pH 5.5. Only buffer and substrate were combined for negative controls. Upon mixing, BC and PASC reactions were immediately placed in a 50 °C water bath. Samples in triplicate were removed at the given time points and placed on dry ice to stop the

reaction. Frozen samples were later placed in a boiling bath for 10 minutes in order to denature the enzyme. It was verified experimentally that boiling does not alter soluble sugar profiles detected by High Performance Liquid Chromatography (HPLC). The remaining substrate was removed using Corning® Spin-X® Centrifuge tube filters and the soluble sugar concentrations were measured using a Shimadzu HPLC system fitted with a Bio-Rad Aminex® HPX-87P analytical column and a refractive index detector. The mobile phase was Milli-Q water at a flow rate of 0.6 ml/min. Sample injection (50 µl volume) was performed by an autosampler installed on the instrument.

BC hydrolysis assays of different concentrations of TfCel48A with constant TfCel9A.

Reactions were prepared in the same manner as for time course assays. 8.3 nM TfCel9A was combined with 0-158 nM TfCel48A and incubated with 1 mg BC for 19 hours. Samples were then processed for HPLC in the same manner as described above.

Sequential BC addition assays:

Assays were conducted under the same conditions as time course assays. Enzymes and substrate were incubated for 20 hours and divided in three groups. One group was boiled for 5 minutes to denature the enzyme and was then processed for HPLC analysis as described above. To the other two groups either 1 mg fresh BC or buffer was added and the reactions were incubated for another 20 hours. Reactions were then boiled and processed for HPLC.

Sequential enzyme addition assays:

Assays were conducted under the same conditions as time course assays. For endpoint sequential assays enzymes and substrate were incubated for 20 hours and boiled for 5 minutes. The corresponding synergistic partner was then added and the reactions were incubated for 2.25 hours. Comparison was made between total soluble sugar produced by enzymes on fresh and

pretreated substrates. In the case of substrate pretreatment the total sugar produced before the addition of the synergistic partner was determined experimentally and subtracted from the total soluble sugar present after digestion by the freshly added enzyme. For time course sequential assays reactions were initially prepared by mixing 83 nM TfCel9A with 1 mg BC. Triplicate samples were collected at the indicated intervals and frozen on dry ice to stop the reaction. After overnight incubation all remaining reactions were boiled for five minutes and divided in two groups. 83 nM TfCel9A was added to one group and 750 nM TfCel48A was added to the other group. Samples from both groups were removed in triplicate at indicated intervals. After overnight incubation remaining reactions containing TfCel48A were boiled for five minutes and 83 nM TfCel9A was added. Samples were removed in triplicate at indicated time intervals. Controls were run to ensure that boiling of the reactions was sufficient to stop all additional activity when incubated at 50 °C. It was also verified experimentally that boiling does not alter HPLC profiles of the soluble sugars or affect the reactivity of BC.

Data analysis.

HPLC data were processed with OriginPro 8. Product identities and concentrations were determined by Gaussian peak fitting, using standard solutions with known concentrations of soluble cellooligosaccharides for reference. Soluble sugar concentrations at time zero were subtracted from all of the subsequently obtained concentrations. Soluble sugar produced upon initial mixing of the enzyme and the substrate is primarily due to the burst activity, as described in (*6*, *7*), whereas the model used here is concerned with the digestion of the more recalcitrant portions of cellulose. *A* and *b* parameter values of time course profiles were determined using the nonlinear least squares fit of Eqn. 6.1.

Kinetic model of cellulose hydrolysis.

The kinetic model described in Chapter 2 was used for this study and is briefly reviewed. The model relies on two parameters to quantify cellulose digestion over time by individual cellulases and their mixtures. It is based on a pseudo zero order Michaelis-Menten kinetic scheme, but replaces an activity constant with a time (digestion) dependent activity coefficient:

$$X = At^b \quad (6.1)$$

where X is percent digestion, A is the total activity of the added enzyme, and b is the "hydrolysis power factor" that quantifies the dependence of activity on time. A and b values are determined by fitting Equation 6.1 to time course data of cellulose digestion. Parameter A is a product of specific activity and the productively bound cellulase concentration, neither of which is directly measurable for cellulases acting on bulk substrates. At low enzyme loads, the value of A is strongly dependent on total added enzyme concentration and follows the Arrhenius relationship with respect to temperature. Parameter b is an intrinsic constant for a given cellulase on a given substrate and is independent of the total added enzyme at low enzyme to substrate ratios. The theoretical limits of b are 0, in which case no products are formed and 1, in which case the activity is constant over time, as is the case in classical kinetics. Thus, the value b is indicative of the cellulases' ability to overcome substrate recalcitrance and it can be used to better understand synergistic interactions between biomass active enzymes. For example, we showed that the addition of a *T. fusca* AA10 enzyme E7 to TfCel48A acting on BC increased the b value from 0.34 to 0.65 (see Chapter 2), which is consistent with the presumed ability of AA10 enzymes to disrupt crystalline regions of cellulose and chitin (*8, 9*).

Results

Time course data of BC hydrolysis by TfCel9A and TfCel48A, alone and together, are shown in Figure 6.1. The very low activity of TfCel48A on crystalline cellulose is reflected by its low *A* and *b* values – 0.77 and 0.35 respectively. On the other hand, relatively high TfCel9A *A* and *b* values of 3.35 and 0.56 respectively reflect its ability to hydrolyze BC well. When the two enzymes are combined together at the same loads as in individual assays, the *b* value of the mixture is 0.55, which is the same as that of TfCel9A alone, and the *A* value is increased by ca. 260% compared to the sum of *A* values measured for the two cellulases alone. Equation 6.1 is typically applicable only for initial range of digestion, which can vary strongly for different enzymes (REF). It is likely that the "drop-off" occurs due to the contribution of additional factors other than substrate recalcitrance to the decline of cellulose digestion rate (e.g. surface erosion and obstacle formation (*10-12*). Importantly, while the "drop-off" point is below 2% digestion for TfCel48A and ca. 12% for TfCel9A, it increases to ca. 35% when the two enzymes are combined.

The fact that the *b* parameter value of the TfCel48A-TfCel9A mixture is the same as that of TfCel9A while the *A* value is significantly increased suggests that the two enzymes acting together are able to digest the BC much faster than either one acting alone, but they are not more effective at overcoming substrate recalcitrance *per se*. BC hydrolysis rate increase for the mixture may be the result of higher activity of either or both enzymes. As a processive enzyme, TfCel9A initially produces cellotetraose (G4), which is later cleaved in solution (by TfCel9A and other cellulases that may be present) to G3, G2, and G1. On the other hand, TfCel48A produces mostly G2 and G3, with only a trace amount of G4, which it is able to hydrolyze only slowly. Therefore, most of the G4 present during the initial stages of BC digestion by the TfCel48A-

TfCel9A mixture is produced by TfCel9A and can be used to compare TfCel9A activity alone and in the mixture. As shown in Figure 6.3a, G4 concentration profile during the initial 14 hours of BC digestion is roughly the same for TfCel9A alone and in combination with TfCel48A. This strongly suggests that the rate of BC hydrolysis by TfCel9A is not significantly increased in the presence of TfCel48A and implies that most of the rate increase observed for the two enzymes acting together is due to additional TfCel48A activity. Further support for this proposition is shown in Figure 6.3b, which demonstrates strong dependence of the total soluble sugar produced with increasing amount of TfCel48A added to a constant amount of TfCel9A in endpoint assays of BC hydrolysis. Interestingly, even though the total soluble sugar produced is strongly dependent on TfCel48A concentrations even at high TfCel48A:TfCel9A ratios, the specific activity of the total added enzyme is highest at the ratio of 0.5.

The main factors that may contribute to the rate decline of enzymatic cellulose digestion at low enzyme loads are substrate recalcitrance, morphological substrate changes, and enzyme deactivation (due to poor thermostability and/or unproductive irreversible binding). To determine whether rate decline of BC digestion by TfCel9A and TfCel48A was caused primarily by substrate or enzyme related changes, we carried out sequential BC addition endpoint assays, shown in Figure 6.2. After BC hydrolysis assays with TfCel48A or TfCel9A alone were incubated overnight, the same amount of fresh BC or buffer was added to the reactions and incubated for the same amount of time. For both TfCel48A and TfCel9A the addition of fresh BC results in much higher activity compared to buffer only addition. Incubation of TfCel9A with fresh and partially digested BC results in the formation of the same amount of additional product as initial digestion, while continued digestion of the original BC with fresh buffer produces only half as much additional product. This suggests that TfCel9A activity is primarily

limited by the substrate-related factors and that the enzyme is not deactivated over the two day period during which the assays were run. Similarly, TfCel48A is able to hydrolyze fresh BC much better than the partially digested BC after overnight incubation, indicating that its activity is also primarily limited by substrate-related factors. However, the activity of TfCel48A on fresh BC results in only 70% additional product formation after the same incubation time, indicating that some fraction of TfCel48A is deactivated during the initial digestion.

Sequential addition of synergistically acting cellulases can provide insight into the roles played by synergistic partners in cooperative substrate digestion (*11, 13, 14*). For this reason we tested the effect of substrate pretreatment by TfCel48A and TfCel9A on each other's activities (Figure 6.3a). Substrate was incubated with either enzyme overnight and the reactions were then boiled to denature the enzymes. The synergistic partner was then added and reactions were again incubated overnight with comparison made to assays prepared with fresh BC. The results indicate that pre-digestion of BC with TfCel9A enhances TfCel48A overnight endpoint activity by ca. 100%. On the other hand, TfCel48A pre-digestion of BC does not enhance activity of TfCel9A under the tested conditions. These observations are consistent with other published data that show enhanced exocellulase activity following substrate pretreatment with an endocellulase, but not vice versa (*14*).

To further understand the role of TfCel48A in its interaction with TfCel9A we carried out more sequential enzyme addition assays such that the initial digestion with TfCel9A was followed up with either fresh addition of TfCel9A or with addition of TfCel48A followed by TfCel9A (Figure 6.3b). The goal was to see whether the intermediate addition of TfCel48A would enhance the activity of the newly added TfCel9A, and this is indeed the case. After overnight incubation with TfCel9A and boiling of the reactions to denature the enzyme addition

of fresh TfCel9A lead to an increase in the extent of BC digestion by less than 2.5 percentage points in four hours. However, when fresh TfCel9A was added following intermediate overnight incubation with TfCel48A, the extent of BC digestion increased by more than 4 percentage points in four hours, indicating that TfCel9A activity was enhanced by ca. 60% by the intermediated addition of TfCel48A.

Discussion

TfCel48A and TfCel9A are important components of the *T. fusca* biomass degrading arsenal and their properties make them incompatible with the classical endo-exo model of synergism. To understand the potentially distinct mode of cooperation between these enzymes we carried out endpoint, time course, and sequential addition BC hydrolysis assays, which provide the following suppositions. When combined, the two enzymes are able to hydrolyze BC much faster than either one alone, but their ability to overcome BC recalcitrance is the same as that of TfCel9A acting alone (Figure 6.1). The hydrolysis rate decline in individual enzyme assays is primarily caused by substrate-related factors for both cellulases, although it also appears that a fraction of TfCel48A is deactivated when the enzyme is acting alone (Figure 6.3). Most of the synergistic product formation is due to the activity of TfCel48A, and the rate of BC hydrolysis by TfCel9A is not apparently enhanced in the presence of TfCel48A (Figure 6.2). Pretreatment of fresh BC by TfCel9A enhances the ability of TfCel48A to hydrolyze the same substrate, but the reverse is not true. However, TfCel48A hydrolysis of BC pre-digested by TfCel9A stimulates further hydrolysis by TfCel9A (Figure 6.4).

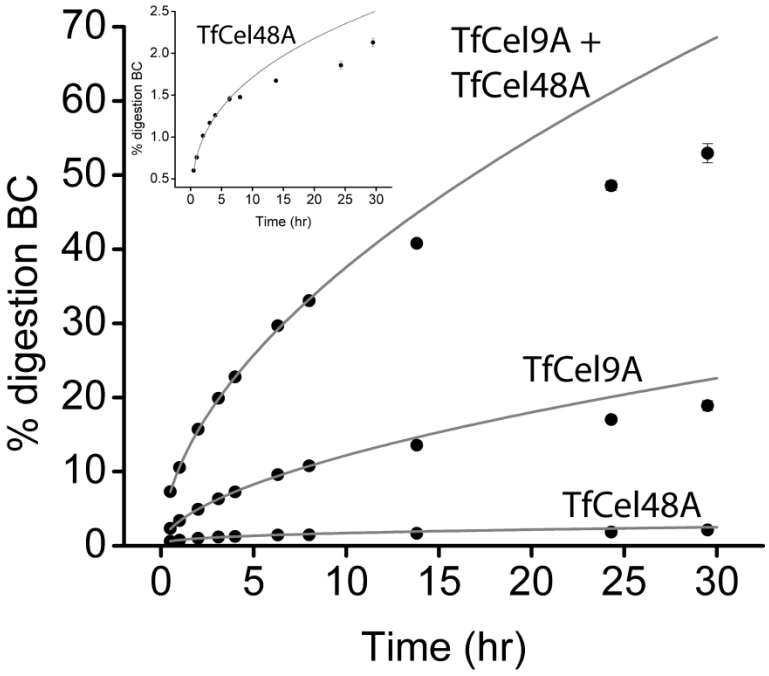
Based on these observations, we propose a model of synergism shown in Figure 6.5. TfCel9A appears to be best adapted for the hydrolysis of fresh BC. CBM3c is necessary for the ability of TfCel9A to hydrolyze BC (*1*) and its relatively flat binding surface may function best

on a uniform substrate surface (*3*, *15*). According to Monte Carlo simulations (*10*) and high resolution imaging before and after digestion (*12, 16*), cellulose hydrolysis by processive cellulases creates an eroded surface, which may no longer be optimal for TfCel9A binding. BC surface erosion may, therefore, explain why TfCel9A activity drops off more rapidly above ca. 12% substrate digestion. On the other hand, the very low activity of TfCel48A on BC (Figure 6.1) suggests that it cannot effectively access the individual chains in this primarily crystalline substrate (*17*). Eroded BC surface formed by TfCel9A activity may provide a much more accessible substrate for TfCel48A, as it creates new chain ends and less tightly bound chains. TfCel48A would preferentially hydrolyze looser chains, replenishing the smooth surface required for efficient TfCel9A activity. Such feedback mechanism would allow for the continuous optimized activity by both cellulases, and this is reflected by their high total activity on BC as well as Equation 6.1 "drop-off" point of digestion well above that observed for either enzyme alone.

A comparable model of synergism was proposed by Valjamae *et al.* (*11*) for *Trichoderma reesei* enzymes TrCel7A (CBH I, a processive exocellulase) and TrCel7B (EG I, a non-processive endocellulase). In their study of BC hydrolysis by the two cellulases, the authors concluded that processive action of TrCel7A creates an eroded surface with obstacles that reduce activity of the enzyme. The modified surface, however, is efficiently hydrolyzed by TrCel7B and allows for sustained TrCel7A activity. TfCel9A and TfCel48A, therefore, appear to play similar roles as TrCel7A and TrCel7B respectively in the model proposed in the above study. It is important to note, however, that the roles of the endo- and exo- cellulases are reversed in the two models, which suggests that *T. fusca* and *T. reesei* may have evolved somewhat different

approaches to biomass digestion. If this is indeed the case, it would be interesting to determine whether such differences apply more generally to cellulolytic bacteria and fungi.

The model of synergism between TfCel48A and TfCel9A proposed here was initiated from the time course data shown in Figure 6.1 by comparing Equation 6.1 parameter values obtained for the individual enzymes and their mixture. Additional assays were then carried out to further develop and verify the TfCel48A-TfCel9A synergism model. We believe that this study provides additional support for the utility of our recently developed kinetic model for the mechanistic studies of enzymatic cellulose digestion.



	A	*b*
48A	0.77±0.02	0.34±0.02
9A	3.35±0.02	0.56±0.01
48A-9A	10.65±0.06	0.55±0.01

Figure 6.1: Time course BC hydrolysis assays of TfCel48A, TfCel9A, and their mixture. TfCel48A time course is magnified in the inset graph. All data points were obtained in triplicate and the error bars represent one standard deviation. Table lists *A* and *b* parameter values obtained by fitting time course data to Equation 6.1.

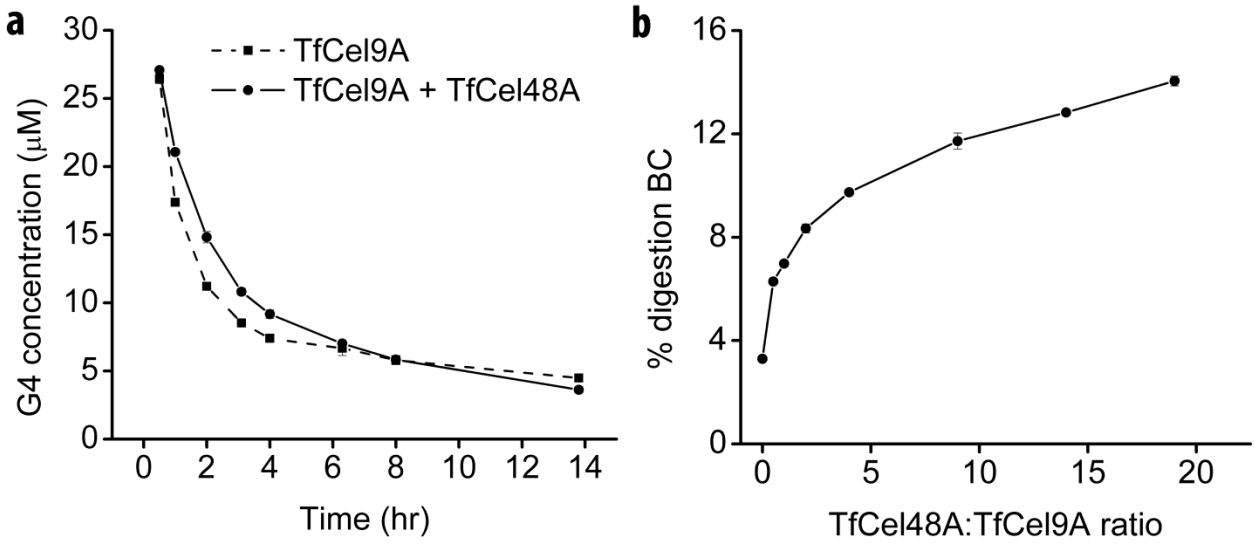


Figure 6.2: TfCel9A and TfCel48A relative activity in the synergistic mixture. a) Cellotetraose (G4) concentration profile for BC hydrolysis by TfCel9A alone and the TfCel9A-TfCel48A mixture. b) Extent of digestion as a function of TfCel48A concentration increase with a constant concentration of TfCel9A (8.3 μM) in BC hydrolysis endpoint assays.

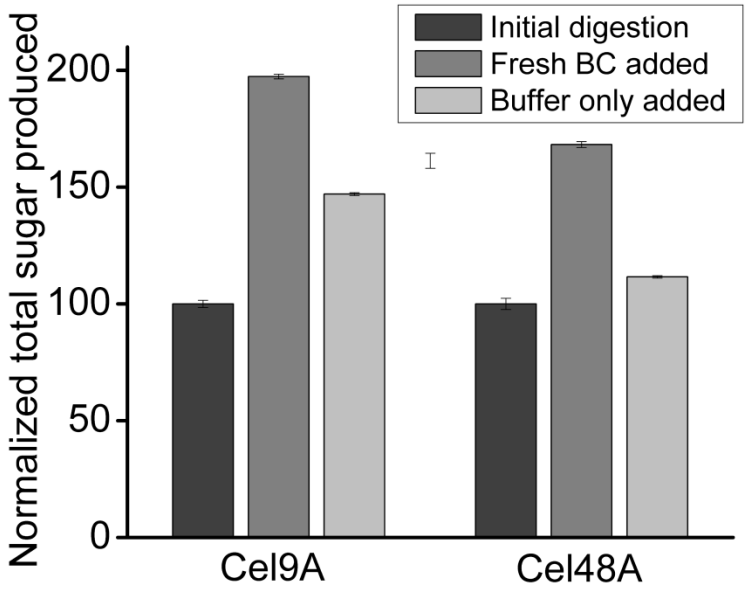


Figure 6.3: Sequential BC addition in endpoint assays. Initial 20-hour BC digestion was followed by the addition of fresh BC or buffer to the reactions. Reactions were then incubated for another 20 hours and normalized soluble sugar measured in the reactions is presented.

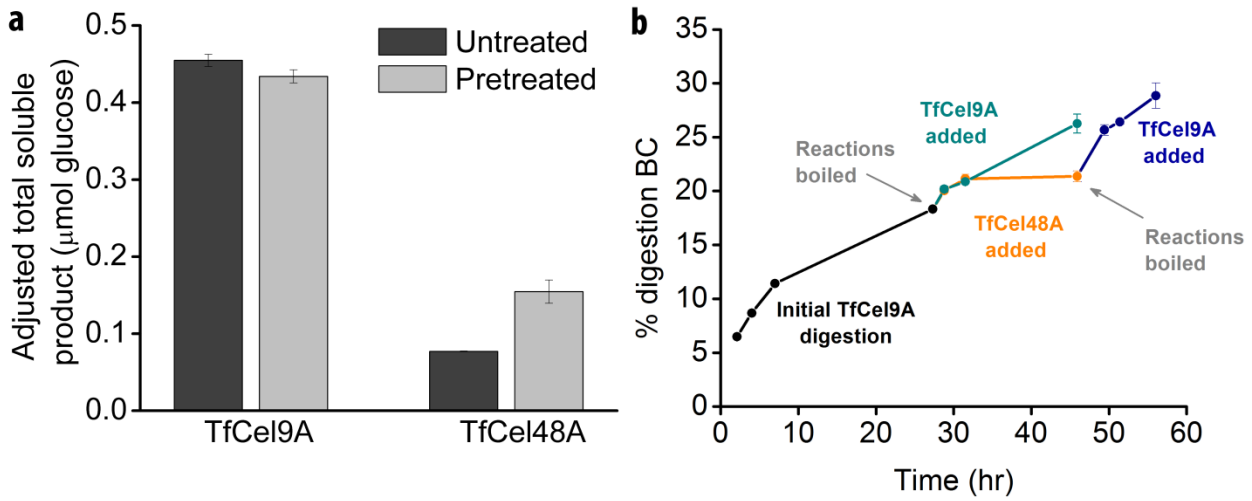


Figure 6.4: Sequential additions of TfCel48A and TfCel9A. a) BC was pretreated with either TfCel48A or TfCel9A for 20 hours and the reactions were boiled to denature the enzyme. The synergistic partner was then added and the reactions were incubated for 2.25 hours. Comparison is provided for activity on untreated and enzyme-pretreated substrate. b) Initial TfCel9A digestion was followed by the addition of TfCel9A with or without an intermediate incubation with TfCel48A. The intermediate addition of TfCel48A enhances activity of the freshly added TfCel9A, as evidenced by the steeper % digestion profile following incubation with TfCel48A.

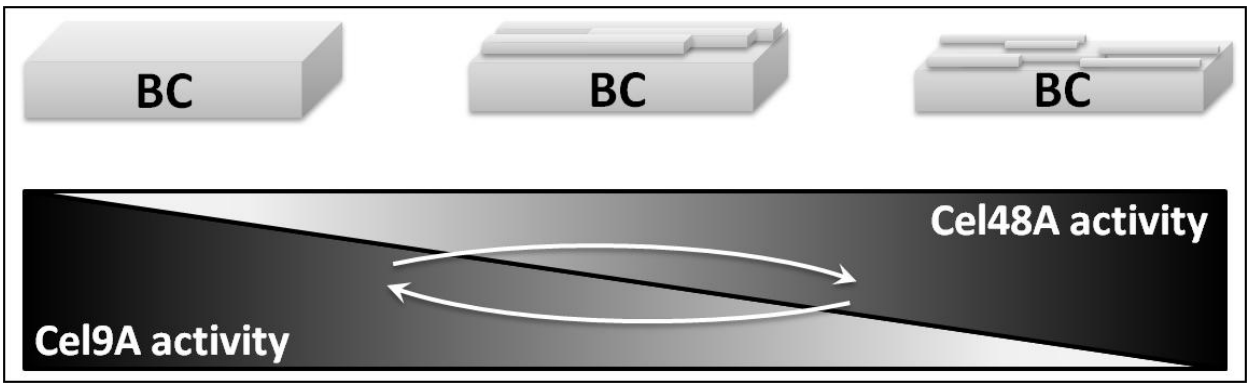


Figure 6.5: Proposed model of synergistic cooperation between TfCel48A and TfCel9A.

References

1. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of Thermomonospora fusca E4 in cellulose hydrolysis, *J Bacteriol 180*, 1709-1714.

2. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects, *Biotechnol Bioeng 42*, 1002-1013.
3. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca, *Nat Struct Biol 4*, 810-818.
4. Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca, *Eur J Biochem 267*, 4988-4997.
5. Spiridonov, N. A., and Wilson, D. B. (1998) Regulation of biosynthesis of individual cellulases in Thermomonospora fusca, *J Bacteriol 180*, 3529-3532.
6. Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K. C., Borch, K., and Westh, P. (2011) A kinetic model for the burst phase of processive cellulases, *Febs J 278*, 1547-1560.
7. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., and Westh, P. (2012) Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A, *J Biol Chem 287*, 18451-18458.
8. Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sorlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides, *Science 330*, 219-222.
9. Horn, S. J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V. G. (2012) Novel enzymes for the degradation of cellulose, *Biotechnol Biofuels 5*, 45.

10. Valjamae, P., Sild, V., Pettersson, G., and Johansson, G. (1998) The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model, *Eur J Biochem 253*, 469-475.

11. Valjamae, P., Sild, V., Nutt, A., Pettersson, G., and Johansson, G. (1999) Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I, *Eur J Biochem 266*, 327-334.

12. Ganner, T., Bubner, P., Eibinger, M., Mayrhofer, C., Plank, H., and Nidetzky, B. (2012) Dissecting and reconstructing synergism: in situ visualization of cooperativity among cellulases, *J Biol Chem 287*, 43215-43222.

13. Jeoh, T., Wilson, D. B., and Walker, L. P. (2006) Effect of cellulase mole fraction and cellulose recalcitrance on synergism in cellulose hydrolysis and binding, *Biotechnol Prog 22*, 270-277.

14. Nidetzky, B., Steiner, W., and Claeyssens, M. (1994) Cellulose hydrolysis by the cellulases from Trichoderma reesei: adsorptions of two cellobiohydrolases, two endocellulases and their core proteins on filter paper and their relation to hydrolysis, *Biochem J 303 (Pt 3)*, 817-823.

15. Oliveira, O. V., Freitas, L. C. G., Straatsma, T. P., and Lins, R. D. (2009) Interaction between the CBM of Cel9A from Thermobifida fusca and cellulose fibers, *J Mol Recognit 22*, 38-45.

16. Chanzy, H., Henrissat, B., Vuong, R., and Schulein, M. (1983) The action of 1,4-beta-D-glucan cellobiohydrolase on *Valonia* cellulose microcrystals. An electron microscopy study., *FEBS Lett 153*, 113-118.

17. Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A., and Johnson, D. K. (2010) Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance, *Biotechnol Biofuels 3*, 10.

Appendix A

Permissions for reuse of published work:

Chapter 1 contains parts of the following publication:

Maxim Kostylev and David Wilson (2012) Synergistic Interactions in Cellulose Hydrolysis. *Biofuels 3*, 61-70.

Journal policy allows reuse of the published work by the authors for non-commercial purposes, such as this dissertation, and no special permission is required.

Chapters 1 and 2 contain parts of the following publication:

Maxim Kostylev and David Wilson (2013) Two-Parameter Kinetic Model Based on a Time-Dependent Activity Coefficient Accurately Describes Enzymatic Cellulose Digestion, *Biochemistry* Article ASAP, DOI 10.1021/bi400358v

In accord with journal policy, the published work can be reused in this dissertation with the following statement included here:

Adapted with permission from *Biochemistry*. Copyright (2013) American Chemical Society.

Chapters 1 and 3 contain parts of the following publication:

Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2012) Determination of the molecular states of the processive endocellulase Thermobifida fusca Cel9A during crystalline cellulose depolymerization, *Biotechnol Bioeng 109*, 295-299.

Written permission to reuse the content for this dissertation was obtained from John Wiley and Sons under the license number 3197120979230.

Chapter 4 contains parts of the following publication:

Kostylev, M., and Wilson, D. B. (2011) Determination of the catalytic base in family 48 glycosyl hydrolases, *Appl Environ Microbiol 77*, 6274-6276.

In accord with journal policy, the published work can be reused in this dissertation with the following statement included here: